INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

NOTE TO USERS

This reproduction is the best copy available.

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

BRILLIANCE, ENERGY AND SIZE IN VOWELS: A CROSS-LINGUISTIC STUDY OF PHONETIC SYMBOLISM

by

STEPHANIE SMOLINSKY

VOLUME I

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2001

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3024835

Copyright 2001 by
Smolinsky, Stephanie

All rights reserved.

UMI®

UMI Microform 3024835
Copyright 2001 by Bell & Howell Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

© 2001

STEPHANIE SMOLINSKY

All Rights Reserved

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Aug 30, 2001
Date

Edward N. Gendel
Chair of Examining Committee

Aug 30, 2001
Date

Charles E Carr
Executive Officer

Dr. Dianne Bradley

Dr. Lawrence Raphael

Professor William Stewart

Supervisory Committee
The City University of New York

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Abstract
Brilliance, Energy, and Size in Vowels:
A Cross-Linguistic Study of Phonetic Symbolism

By Stephanie Smolinsky

Adviser: Professor Edward Bendix

The dissertation explores the question of *segmental phonetic symbolism*—the attribution of qualities such as brightness and largeness to speech sounds. An experiment investigates attributions for the point vowels [i], [a] and [u], and the mid central vowel [ʌ], paired stimuli being specified in [hV_1, hV_2] form. A questionnaire asking *Which of the two stimuli is more X?* for adjectives *bright* and *dark*, *active* and *passive*, *large* and *small* (representing brilliance, energy, and size dimensions, respectively) elicited judgments from speakers of American English, Korean, Japanese and Spanish.

These judgment data are analyzed in terms of three basic questions. First is whether evidence exists of a consistent phonosymbolic response, the claim here turning on the extent of statistically significant concurrence about the stimulus preferred in any pair-comparison, within language groups. This is answered with a strong affirmative. Second is whether phonosymbolism should be ascribed a universal or language-specific basis, as determined by agreements/disagreements across language groups about particular stimulus rankings (e.g., brightest to least bright). Findings suggest a predominant universal basis, tuned in special circumstances by a language- or culture-specific influence. Third is which vowel-features (phonetic or phonemic) are likely to have underpinned the stimulus rankings. The dissertation proposes several features plausibly triggering phonosymbolism, and suggests mechanisms for their joint application (feature harmony, equivalence, and conflict).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To my parents, Gwen and Harry Smolins,
whose love, support and interest kept me going
when I might otherwise have given up.

In memory of Paul Eden, American English Subject Number Seventy, still much missed.
Of Nellie Finn, whose generosity financed a big piece of this project.
Of David H. Rosenthal, original catalyst, had he but known it.
Of Clinton Wilding Smith, who loved language as much as dancing and gardening.
Of gentle Andrew Tuffin, Americanophile.
Of Jeff Vogel, whose ear, mind, voice, should have been here to consult with.
Of Paul Zmiewski, whose advice on work and health I still try to follow.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Acknowlegments

Many people have helped me during the lengthy production of this work.. My first thanks go to the members of my Committeee. To Dr Edward Bendix, my Chair, for his gentle patience and supportiveness. And for opening the door to linguistics, many years ago, in a classroom in Hunter College.To Dr. Dianne Bradley, for showing me how to pay scrupulous attention to detail, yet also how to let go of details and not be afraid of generalizing: she has educated me more than she will ever know. The beautiful graphics (ranking diagrams of the vowels) were created by her. To Dr. Lawrence Raphael, above all for his cheerful willingness to go over certain acoustic-articulatory connections until they finally sank in. To Professor William Stewart, always available to discuss ideas, who suggested wonderful, unexpected books, like Darwin's, that turned out to be just what I needed.

To Dr. John Roy, who in a café in Fort Tryon Park gave me the original inspiration for the questionnaire. To Dr. Charles Cairns, who provided me with most of my hundred American English-speaking subjects, and who gave me my first chance to air my ideas in public, in his class. To Dr. Winifred Strange, generous with her time and with Japanese and Spanish acoustic data. To my back-translators, Jonathan Bidol (for Korean) and Takako Katsuno (for Japanese), who also answered many questions. To Amy Mandelker for her interest in the early stages of the project, and to Vicky Tartter, for opening my eyes to the *gestaltisch* way that we perceive vowels. To Dr. Ricardo Otheguy, who helped me think through some of the Spanish problems. To Stan Sokolov, without whose long-term computer advice, help and telephone pateince the project would never have gotten off the ground, or stayed in the air. And to Julia Dizenko and Arkasha Kraportov, who rescued me with a hug and the loan of their printer at the last minute.

To my family: my mother who listened so thoughtfully and gave excellent advice, that of a great teacher as well as a loving Mum. To my father, who always told me he was proud of me for keeping on, and that he would read my dissertation when I was through (if he can manage, *mazel tof*: his subject is physics, not linguistics). To my niece Gracie for believing in me and saying so.

To my Zen teacher Anna Christensen, a fountain of compassion and timely kicks, reminding me that this also has been practice. And to all the members of our *Just This* Sangha, cheering me on.

To my dear friends...to Eugene and Natalia Dizenko, to Tim McDonnell and his Assistants Jack and Pierre, to Christel Andino, to Sarah MacMillan, to Susan Seigrist, to Barb Leclerc and the other members of the Ladies' Dinner: Mary Ann McClure, Andrea Siegel, Lee Che Leong and Tisha Ulmer. To Anona Evans, who has always taken my thoughts seriously, to Ivy Sichel who nudged me to get on with it, at just the right moment. To John Pittman who assured me that my argumentation-writing methods were quite normal. To my cat Snippins, who explained Darwin's insights on the nature of metaphor. To Linda Pelc, who found me scores of subjects, to Charlotte Balfour who found me my last Japanese subject and to Luis Miguel Capriles, who *was* my last Spanish-speaking subject. To Margaret Magnus, the first person with whom I have dared to share certain ideas, who has spend many years thinking about phonetic symbolism.

And to Laura Greenberg, my compañera in teaching, writing, swimming...my coach in dissertation-finishing...and so many other grownup skills. Thankyou.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CONTENTS

Volume I

Abstract....iv
Dedication....v
Acknowledgments....vi
List of Tables....x
List of Figures....xii

CHAPTERS

1 Introduction....1
2 Phonetic symbolism: Plato to 1933....9
3 Lexical and poetical studies, 1934 to the present....58
3.1 General introduction....58
3.2 Studies....60
4 Fittingness and synesthesia studies, 1934 to the present....126
5 Two lines of psycholinguistic experiment, 1934 to the present....184
5.1 Word-matching experiments....185
5.2 Minimal pair experiments....214
6 Experimental aims, method and interpretational approach....236
6.1 Experimental aims....236
6.2 Experimental method....237
6.3 Additional approaches to interpreting the data....258
7 Analysis and interpretation of the American English data....261
7.1 Phonetic/phonological background of AE-speaking subjects....261
7.1.1 Articulatory details....261
7.1.2 Acoustic details....264
7.1.3 Vowel scales of two important predecessors....268
7.2 Analysis of the American English data....270
7.2.1 Brillliance dimension....274
7.2.2 Energy dimension....281
7.2.3 Size dimension....290

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

7.2.4 Concluding remarks on AE data & phonosymbolic profiles of stimulus vowels according to AE-speaking subjects 300

Volume II

8 Analysis and interpretation of the Korean data.....305
8.1. The phonetic/phonological background of Korean subjects.....305
8.1.1 Articulatory details.....305
8.1.2 Acoustic details.....311
8.2. Analysis of the Korean data.....320
8.2.1 Brillliance dimension.....323
8.2.2 Energy dimension.....327
8.2.3 Size dimension.....332
8.2.4 Concluding remarks on the Korean data & phonosymbolic profiles of the stimulus vowels according to Korean subjects...338
9 Analysis and interpretation of the Japanese data.....348
9.1 The phonetic/phonological background of Japanese subjects.....348
9.1.1 Articulatory details.....348
9.1.2 Vowel devoicing.....360
9.1.3 Acoustic details.....365
9.2 Analysis of the Japanese data.....386
9.2.1 Brillliance dimension.....395
9.2.2 Energy dimension.....402
9.2.3 Size dimension.....409
9.2.4 Concluding remarks on the Japanese data & phonosymbolic profiles of the stimulus vowels according to Japanese subjects..... 425
10 Analysis and interpretation of the Spanish data.....430
10.1 The phonetic/phonological background of Spanish-speaking subjects
10.1.1 Articulatory details.....430
10.1.2 Acoustic details.....463

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

10.2 Analysis of the Spanish data.....480
10.2.1 Brillliance dimension.....487
10.2.2 Energy dimension.....490
10.2.3 Size dimension.....494
10.2.4 Concluding remarks on the Spanish data & phonosymbolic profiles of the stimulus vowels according to Spanish-speaking subjects.....504
11 Phonetic symbolism across four languages.....510
11.1 The existence of phonetic symbolism.....511
11.2 Universality vs. language-specifity in phonetic symbolism.....513
11.2.1 Consensus distribution data & universality vs language-specificity.....514
11.2.2 Ranking data & universality vs. language-specifity.....529
11.3 Cross-linguistic personalities of the stimulus vowels....557
12. Conclusions and suggestions for further research.....559
12.1 Phonosymbolic triggers, compared with those proposed by Sapir & Newman.....559
12.2 Summary & suggestions for further research.....565
Appendices
1 Questionnaire materials.....571
2 ESL subjects' self-reported experience in learning/speaking English.....579
3 Frequency totals and ranking index values.....580
4 Data patterns for all language groups.....583
5 Frequency matrices for all language groups.....587
6 Fónagy's thinking on phonetic symbolism.....591

References 617

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF TABLES

Table

7.1 Formant frequencies (Hz) and intensities (dB) for male AE speakers.....265

7.2 Ranking index ranges for two semantic dimensions: AE speakers.....283

7.3 Ranking index ranges for three semantic dimensions: AE speakers.....292

7.4 Observed vs predicted size rankings: AE speakers.....294

8.1 Formant frequencies (Hz) for male Korean speakers.....313

8.2 Formant frequencies (Hz) for male AE speakers.....313

8.3 F_0 values for male Korean speakers.....319

8.4 F_0 values for male AE speakers.....319

8.5 Ranking index ranges for two semantic dimensions: Korean speakers.....329

8.6 Ranking index ranges for three semantic dimensions: Korean speakers.....334

8.7 Observed vs predicted size rankings: Korean speakers.....336

9.1 Japanese female F_1 & F_2 values (Hz).....367

9.2 Japanese male F_1 values (Hz).....369

9.3 Japanese male F_2 values (Hz).....370

9.4 Comparison of Japanese & AE female F_1 values (Hz).....373

9.5 Comparison of Japanese & AE female F_2 values (Hz).....373

9.6 Comparison of Japanese & AE male F_1 values (Hz).....374

9.7 Comparison of Japanese & AE male F_2 values (Hz).....375

9.8 F_0 through for F_2 values (Hz) for point vowels: one female Japanese speaker.....377

9.9 F_1 through for F_3 values (Hz) for point vowels: male & female Japanese speakers.....378

9.10 F_1 through for F_3 values (Hz) for point vowels: one male Japanese speaker.....378

9.11 F_1 through for F_3 values (Hz) for point vowels: four male Japanese speakers.....378

9.12 F_1 & F_2 values (Hz) for point vowels: male Japanese speakers (#''s unspecified).....379

9.13 F_1 & F_2 values (Hz) for point vowels: five male & five female Japanese speakers.....380

9.15 Ratio of relative durational values of point vowels: three Japanese speakers.....381

9.16 Values of avg. point vowel durations (ms): one female Japanese speaker.....381

9.17 Values of avg. one-mora & two-morae point vowel durations: four male Japanese.....382

9.18 Values of avg. point vowel amplitudes (dB): one female Japanese speaker.....383

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

9.19 F_0 values for point vowels: one female Japanese speaker....384
9.20 F_0 values for point vowels: five male Japanese speakers....384
9.21 AE, Korean & Japanese response range & discrepancy val's: three sem. dimensions...389
9.22 Korean & Japanese response range & discrepancy val's: three sem dimensions....411
9.23 Observed vs predicted size rankings: Japanese speakers....417
10.1 Nationalities of the Spanish-speaking subjects....458
10.2 Comparison of male F_1 values (Hz) : Spanish speakers....468
10.3 Comparison of male F_2 values (Hz): Spanish speakers....469
10.4 Comparison of female F_1 values (Hz): Spanish & AE speakers....470
10.5 Comparison of female F_2 values (Hz): Spanish & AE speakers....471
10.6 Comparison of male F_1 values (Hz): Spanish & AE speakers....472
10.7 Comparison of male F_2 values (Hz): Spanish & AE speakers....472
10.8 Male F_1 through F_3 values (Hz): two Panamanian Spanish dialects....475
10.9 Male F_1 through F_2 values (Hz): Puerto Rican Spanish and AE....476
10.10 Male durational values (ms): vowels of two Panamanian Spanish dialects....477
10.11 Male F_0 values (Hz): vowels of two Panamanian Spanish dialects....479
10.12 AE, Ko., Ja. & Spanish response range & discrepancy val's: three sem. dimensions...482
10.13 Adjectival terms: Spanish translations, cognates & mistranslations....484
10.14 Spanish response range & discrepancy val's: two vitality dimensions....491
10.15 Spanish response range & discrepancy val's: three semantic dimensions....496
10.16 Observed vs predicted size rankings: Spanish speakers....502
11.1 Total number of items evoking consensus....511
11.2 Non-consensus distribution, by stimulus pair....515
11.3 Instances of non-consensus requiring interpretation....516
11.4 Stimulus rank orders and rank steps, for salient scales....531

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF FIGURES

Figure

2.1 Tolman's tone-color scale....7
2.2 Newman's brilliance and size scales....44
7.1 Early vowel scales....270
7.2 Data patterns (AE speakers)....271
7.3 Brilliance dimension data arrays (AE speakers)....274
7.4 Energy dimension data arrays (AE speakers)....281
7.5 Size dimension data arrays (AE speakers)....290
7.6 Preference strength in antonymic scales (AE speakers)....300
8.1 Data patterns (Korean speakers)....320
8.2 Brilliance dimension data arrays (Korean speakers)....323
8.3 Energy dimension data arrays (Korean speakers)....327
8.4 Size dimension data arrays (Korean speakers)....332
8.5 Preference strength in antonymic scales (Korean speakers)....343
9.1 Data patterns (Japanese speakers)....386
9.2 Totals of Japanese translations & mistranslations of adjectival terms....390
9.3 Brilliance dimension data arrays (Japanese speakers)....395
9.4 Energy dimension data arrays (Japanese speakers)....402
9.5 Size dimension data arrays (Japanese speakers)....409
9.6 Preference strength in antonymic scales (Japanese speakers)....426
10.1 Data patterns (Spanish speakers)....480
10.2 Brilliance dimension data arrays (Spanish speakers)....488
10.3 Energy dimension data arrays (Spanish speakers)....490
10.4 Size dimension data arrays (Spanish speakers)....494
10.5 Preference strength in antonymic scales (Spanish speakers)....505
11.1 Preference strength in antonymic scales across four languages....530

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 1

INTRODUCTION

"We are really dealing with a measurably independent [linguistic] factor that for want of a better term may be called *phonetic symbolism*." So wrote Edward Sapir (1929: 283), in an experimental report which was to usher in decades of investigation and controversy, still ongoing. If consistent symbolism is attached to speech sounds (as in Sapir's conclusion that [a] is experienced as large and [i] as small), then one of the keystones of modern linguistics, de Saussure's (1959) concept of the arbitrariness of the sign, is under attack. In some periods, phonetic symbolism has been considered a crucial psycholinguistic issue; in others, the fascination has apparently waned and the whole topic been set aside as eccentric and fly-by-night, not really germane to linguistics. But interest has again and again regenerated: evidence that the whole argument over phonetic symbolism—its existence, scope and nature, how it can be validly tested and how results should be interpreted—is properly part of the scientific investigation of language.

1.1 Choosing a label. In the literature on this subject, two terms are used, often as if they were interchangeable: *phonetic symbolism* and *sound symbolism*. Yet these are not synonymous labels. Whichever term is used makes an implicit claim about the nature of what is being studied. *Sound symbolism* is the broader term, referring to symbolisms not only attaching to speech sounds, but to all sounds. It names that general type of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

impressionistic attribution that we make when we say that a trumpet has a "bright" sound, that the cheeping of a sparrow is "thin" or "high", that the crash of ocean waves is "deep" or "huge", or the rushing sound of an air conditioner's fan is "soft" (i.e., that though it may be quite loud, it does not give an impression of hardness but the reverse). Metaphoric attributions of this type are normal in everyday life: they seem indeed to be built into the language (a "high" wail, a "low" growl, a "sharp" whistle, a "soft" coo, and so on).[1]

However, what to make of a phonetic case, such as the traditional labelling of the velarized and non-velarized allophones of /l/ as "dark [ɫ]" and "clear [l]"? The problem with using the term *sound symbolism* to cover strictly phonetic/phonemic material along with air conditioners and bird calls is that it implies that the attributions that we make are (at least humanly speaking) domain-universal.[2] In other words, there is no room left for the possibility that when we call a speech sound "bright" or "small" there could be *something else* operating besides that general faculty that hears the trombone as dark or the harp and piccolo as bright. And this possible 'something else' is the identity of the speech sound in the language of the speaker as part of a total linguistic pattern, with its unique phonetic/phonemic, morphological and lexical role to play.

---

[1] They are, of course, not used only for qualifying sounds, but for a wide range of sense experiences and abstract concepts, as in "hard *times*," "an icy *stare*," "a warm *smile*," "a bitter *thought*," "sharp *smell*," "strong *taste*," and so on.

[2] But see Ohala (1994) and Morton (1994) for the argument that universality of sound symbolism extends down into the animal kingdom.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Yet on the other hand, if we use *sound symbolism* for attributions to trombones and air conditioners and *phonetic symbolism* for attributions to speech sounds, then the opposite claim is being made: we assert that there is an important intrinsic difference between linguistic and nonlinguistic noise. It then follows that attributions to linguistic noises might vary from language to language, since each phone plays a unique part in its own language.

Let us consider one case here as an illustration: that of the speech sound /ɪ/, a phoneme of English. In the total phonemic configuration of the language, it is a member of the following classes: a) sonorants, b) vowels, c) short (or lax or nondiphthongized) vowels d) high vowels e) front vowels, f) unrounded vowels, and it is also g) the lax member of of the /ɪ/—/iy/ pair. Consequently, /ɪ/, deriving its identity from these memberships, has various articulatory and acoustic features which are potentially available for symbolic use in the way that the feature of velarization of [ɬ] is metaphorically referred to as "dark" and its absence, "bright" or "clear."[3] Thus, if we were to contrast the /ɪ/-/iy/ pair and ask a number of English speakers which seemed to be the smaller vowel, we might expect that a significant number would answer /ɪ/, since its qualities of shortness/laxness/non-dipthongization compared to /iy/ could translate symbolically as "smaller."[4]

---

[3] It could, of course, be the other way round; what is assumed here is that there is a semantic opposition in which the marked member (the velarized [ɬ]) is assigned the feature "dark". In contrast, the other member of the pair then gets assigned the opposite feature without perhaps having to be intrinsically particularly "bright". This follows Saussurean principles of opposition (de Saussure, 1959) and also Darwin's Principle of Antithesis (see Chapter 2).

[4] For experimental corroboration of this viewpoint, see Newman (1933: 68, emphasis added): "The position of [iy] above [ɪ] on the present magnitude scale runs counter to both the articulatory and frequency sequence. The conclusion can hardly be avoided that [iy] is symbolically larger than [ɪ] *because the length of the former vowel makes it 'sound' larger.*"

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Morphologically speaking, in minimal triad verb paradigms such as *begin-began-begun*, *ring-rang-rung*, *swim-swam-swum*, /ɪ/ indicates the unmarked or present tense form of the verb. Lewis (1986: 69) argues that what he calls the "unmarked" form is in semantic opposition to the past or "remote" form, i.e., that the present tense carries a chronologically proximal meaning. In the demonstrative pronoun paradigm *this-these-that-those* /ɪ/ indicates the singular and proximal form (combined with variation in final consonants). Together with the initial consonant variation, in the adverbial pair *here-there* /ɪ/ also indicates the proximal form in some dialects.[5] In other words, we can see a morphologically consistent association between the "close" or "intimate" paradigm member (present tense of the verb, proximal form of demonstrative and adverbial) and the phoneme /ɪ/. This suggests that there is a semantic feature of the segment which might also be available for symbolic use: if we were to contrast /ɪ/ with /æ/ and ask English speakers which of the two vowels was felt to be "nearer", we might expect the answer /ɪ/.

Although this particular phoneme has played no part in the research to follow, it is considerations such as in the above brief description of /ɪ/ which have led me to choose the narrower term *phonetic symbolism* for speech sound attributions. The claim, therefore, is that the specialized nature of phonetic/phonemic material justifies the use of a specialized term. Consequently, when Sapir labels the phenomenon he is examining "phonetic symbolism", his example will be followed wherever those symbolisms derived from phonetic/phonemic objects are meant. Where phonetic symbolism clearly accords with the

---

[5] For support for this view of /ɪ/ as an adverbial and pronominal proximal marker, see Jespersen (1922: 402-3) and Tanz (1971).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

more general phenomenon of sound symbolism, this will be indicated. Indeed the extent to which that accord exists, whether it includes the animal kingdom, or is an isolatedly human, linguistic universal, and to what degree it is limited by language-specific attributions—these are some of the great debates about phonetic symbolism. It is to be hoped that the research described later in this dissertation will shed some light on the question of universality.

An embarrassment of riches. The literature on phonetic symbolism is vast and unwieldy; in one direction, it shades off from pure linguistics into studies of gestalt psychology, cognition and synesthesia, while in another it steps into the work of the Russian Formalists, semiotics, stylisitics and poetics. My knowledge of languages (limited to Spanish and French) has constrained the survey that follows. Certain, apparently seminal, works have never been translated out of their original Russian or German, and therefore will not treated here. Furthermore, there exists untranslated work in Japanese,[6] and perhaps in Korean also, focusing on those separate syntactic classes of phonosymbolic words, mostly adverbial and adjectival forms, known as "expressives," or "attitudinals".[7] Although of obvious relevance, such work has not been available to me. On the other hand, the good luck of knowing French introduced me to Maurice Grammont (1933), who

---

[6] For a frustrating experience (unless one knows Japanese), see Hamano's (1986) bibliography.

[7] Korean and Japanese are two amongst the world's languages (which include some African and other South East Asian languages) which have expressives. New coinages of expressives are always understood by speakers; thus, not only is a high—and accessible—degree of systematization of phonetic symbolism to be expected in languages which have them, but there is a strong likelihood of native-language linguists' exploring this area.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

has written in authoritative detail on phonetic symbolism, and particularly on vowels.[8] It has also allowed me to read the work of Fónagy, whose *Métaphore en phonétique* (1979) and *La vive voix* (1983), unavailable in English, trace highly plausible connections between the infant's experience of language and its adult residue in phonetic symbolism .

It should be added that to read anything of literature on phonetic symbolism is to read something of the cultural history of the last hundred and fifty years, and more especially of the last seventy. Take, for example, Sapir's (1927) original article on the two dimensions of language: the *expressive* (underlying, emotional, primitive, iconic) versus the *referential* (immediately evident, detached, conventional, arbitrary). This interest in a primitive linguistic dimension—and the intent to explore it, rather than hide behind Saussurean dogmas of arbitrariness—can be compared to Lawrence's exploration of the primitive in literature, Stravinsky's in music, Picasso's in art, or the work of Freud, Jung and Groddeck in psychology. Again, much of the more recent debate over the universal versus the language-specific in phonetic symbolism is part of the argument between the behavorists (phonetic symbolism as a set of conditioned reactions due to learning a specific language's vocabulary), the gestaltists (phonetic symbolism as part of a central sensory mediating process which feeds into one's early experience of language) and the Chomskyites (phonetic symbolism as evidence of an underlying universal linguistic

[8] To my knowledge, not only has Grammont's (1933) *Traité de phonétique* never been translated into English, but I have never yet in the literature seen a single quotation from this work, either in English or French, though it is occasionally to be found in the bibliographies of francophone writers on phonetic symbolism.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

structure). I would argue that all these viewpoints have something to offer, and that the best approach involves a synthesis.

The chapters. Chapter 2 will survey writing and research on phonetic symbolism from Plato's *Cratylus* until 1933, which, it can be claimed, is a watershed year for this topic, with a large number of important publications appearing then. The three subsequent chapters will divide studies in phonetic symbolism into three separate strands, following them up to the present. Chapter 3 focuses on lexical and literary-poetical studies, Chapter 4 on studies of fittingness and synesthesia and Chapter 5 on two designs of psycholinguistic studies: those using word-matching and those using minimal pairs.

The research which will be described in Chapters 6 through 11 was an experiment designed to investigate a very small area of phonetic symbolism: symbolisms attaching to isolated vowel sounds. Many phonetic symbolism studies have focused on words and their 'fittingness,' the sense that a word is particularly appropriate to its referent. The problem is that without an already-established sense of the symbolisms attaching to their component sounds, words are bewilderingly complex as objects of study, and a number of interacting factors may be triggering the sense of their fittingness: to name three, effects of contiguous vowels on consonants and vice versa, effects of stress and syllable pattern, and chains of association to other words. In order to avoid confusion through premature involvement with these issues, this experiment sought to establish the pre-lexical

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

'personality' of the vowels as an essential first step in studying the whole phenomenon.[9] As well as American English speakers, the study included three subject groups who were speakers of other languages (Korean, Japanese and Spanish). In this way, the four subject-groups' symbolic attributions to the vowels could be compared, and the question of whether phonetic symbolism is a language-specific or universal phenomenon could be explored. Chapter 6 describes the design and implementation of the experiment, while Chapters 7, 8, 9 and 10 separately report on the findings for each of these language groups. Chapter 11 brings all the findings together, attempting to draw some general conclusions. Based on these, Chapter 12 suggests some modifications to the earlier theories of Sapir (1927, 1929) and Newman (1933), and outlines some areas of fruitful further research.

[9] A projected, later experimental step could then be the study of vowels combined with consonants in non-lexical, nonsense items (i.e. studying the effects of consonants on 'known' vowels and vice versa, but still screening out lexical influence as far as possible). Only after both of these preliminary stages had been accomplished would it be worth proceeding to the study of actual words.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 2

PHONETIC SYMBOLISM FROM PLATO TO 1933

In pre-Sapirian writing on phonetic symbolism, two main preoccupations can be distinguished. One is the great debate about the arbitrariness versus the fittingness of words, and whether the form-meaning connection is random or determined. Since the *Cours génerale de linguistique* was first published (de Saussure, 1915), a fact central to its author and to the linguists that have followed him (including Sapir) is that "The bond between the signifier and the signified is arbitrary" (de Saussure 1959: 67). In other words, the same signified (referent)—*dog* for example—may be randomly symbolized by many different signifiers (words): "dog", "chien", "Hund", "gos", and so on. What connects the animal and the word we use for it in our own language is a linguistic convention that has nothing to do with its sound-shape. This central observation of de Saussure's makes an immediate appeal to common sense. It liberates the linguistic observer from a kind of magical thinking by setting up an impenetrable barrier between the phonetics/phonology of a language and its semantics, breached only in the case of onomatopoeia—and even then, much less than we might think, since any imitation is conditioned by the phonetic/phonological, morphological and lexical structure of the language of the imitator.[1]

[1] See de Saussure (1959: 69), who makes this point.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The argument about fittingness includes phonetic symbolism in the larger concern of what may be termed *lexical* symbolism,[2] which in the West[3] began with Plato However, the fittingness debate (which is as much lexical as phonetic/phonemic[4]) is intertwined with discussion of a related issue, that of the potentially mimetic nature of segments, primarily for literal, this-speech-sound-equals-that-referent-sound use in onomatopoeia, but also in more abstract applications (as in the case of clear [l] versus dark [ɬ]). Given the dual nature of phonetic segments—articulatory and acoustic—the discussion may focus on either aspect or link them together. In any case, this is where we find the true beginnings of an exploration of phonetic symbolism.

Debate over both words and sounds may be found in *Cratylus*, and since familiarity with Plato's thought has been one of the distinguishing marks of an educated person, we can therefore say that the questions of lexical fittingness and phonetic mimesis must have entered Western consciousness around 300 B.C. *Cratylus* opens with a question about the essentiality of the names of things: "Our friend Cratylus has been arguing about names. He says that they are natural and not conventional—not a portion of the human voice which men agree to use—but there is a truth

---

[2] Lexical symbolism and phonetic symbolism are not the same, though it seems hardly possible that the former could exist without some considerable contribution from the latter. It has been argued in Chapter 1 that each phoneme, via its class memberships, makes a number of features potentially available for it to be used symbolically. Words of their nature are composed of more than one phoneme. Hence, any marked fittingness of a word to its referent may be due to a number of segmental and prosodic factors; lexical associations constitute an additional set of influences.

[3] Apparently, it was also a focus of contemporary Chinese philosophical argument: "About the time Plato was writing *Cratylus*...discussion of this question was also going on in China, especially among the Confucianists. The Chinese in general came to a different conclusion than Plato's *Cratylus*" (Lapolla 1994: 130).

[4] In these literature review chapters, where the focus is on the total sound system of a language (i.e. the phonetic *together with* the phonological), this will be referred to hereafter by the wider of the two terms as "phonetic." "Phonological" will be used purely to refer specifically to phonemic distinctions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and correctness in them which is the same for Hellenes as for barbarians. Whereupon I ask him whether his own name of Cratylus is a true name or not, and he answers yes" (*Cratylus*: 383 b).

The discussion first appears to support the idea that there is some principle in language at work to ensure that things are named appropriately, and that even language-to-language variation in naming the same object (one of de Saussure's main pieces of evidence in support of arbitrariness) can be reconciled with a principle of fittingness.[5] A number of names of humans, places, Gods, heroes and qualities are now examined in terms of their appropriateness to their referents; judgments of good or bad[6] fit are made on the basis of literal meaning, supported by morphology: "The irreligious son of a religious father...should not be called Theophilus (beloved of God) or Mnesitheus (mindful of God)" (*Cratylus*: 394 e) and of apt metaphorical extension: "He is rightly called Pelops who sees what is near only." (*Cratylus*: 395 c).

However, Socrates then observes that close examination of word meanings will gradually narrow down into examination of "primary elements," i.e., component segments, and that this type of examination will involve "some new method" (*Cratylus*: 421 e--422 b) The new method of judging fittingness turns out to be *phonetic*, and the way Socrates arrives at it is also of interest, since he starts by apparently anticipating certain arguments of Darwin's, on the gestural origin of language (on which Sapir was to base his own ideas and investigations twentythree hundred years

---

[5] See *Cratylus*: 387 d, 388 d-e, and 390 b. For the moral aspect of correct naming, see also *Cratylus*: 390 e.

[6] Errors of fittingness are explained as due to the foreign origins of certain words or to linguistic change: "The original names have been long ago buried and disguised by people sticking on and stripping off letters for the sake of euphony, and twisting and bedizening them in all sorts of ways, and time too may have had a share in the change." (*Cratylus*: 414 c-d)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

later).[7] But then these same arguments are dismissed; Socrates cannot conclude the process is so literal that a name can simply be described as "A vocal imitation of that which the vocal imitator names or imitates,"(*Cratylus*: 423 b-c) since if it were, the conceptual element of the name would disappear, swallowed up by onomatopoea: "We shall be obliged to admit that the people who imitate sheep or cocks or other animals, name that which they imitate." (*Cratylus*: 423 c-d)

Instead, he asserts that the sounds of words are related to the essential qualities of their referents and are analyzable into their constituent phones: "Imitation of the essence is made by syllables and letters"[8] (*Cratylus*: 424 b-c) "That objects should be imitated in letters and syllables, and so find expression, may appear ridiculous... but it cannot be avoided. There is no better principle to which we can look for the truth of first names. Deprived of this, we must have recourse to divine help." (*Cratylus*: 425 d-e) Then various "letters" (i.e., sounds) are analyzed in terms of their meaning.[9] Disappointingly for those interested in phonetic symbolism, the argument of *Cratylus* concludes with a warning against expecting too much symmetry between language and the things it

---

[7] These are worth quoting in detail: "Suppose we had no voice or tongue, and wanted to communicate with one another. Should we not, like the deaf and dumb, make signs with the hands and the head and the rest of the body?... We should imitate the nature of the thing: the elevation of our hands to heaven would mean lightness and upwardness; heaviness and downwardness would be expressed by letting them drop to the ground...And when we want to express ourselves with the voice, or tongue, or mouth, their expression is simply the imitation of that which we want to express." (*Cratylus*: 422 e--423 b-c)

[8] "Letter" here means both the character that represents the sound *and* the sound, as it does in the subsequent quotations from Bacon (1631) and Guest (1838).

[9] To give their flavor, these descriptions are again worth quoting in some detail: "The letter *rho* appears to me to be the general instrument expressing all motion... the tongue is most agitated and least at rest in the pronunciation of this letter....There is another class of letters, *phi*, *psi*, *sigma* and *zeta*, of which the pronunciation is accompanied by a great expenditure of breath; these are used in imitation of such notions as *psuchron* (shivering), *zeon* (seething), *seiesthai* (to be shaken).... and are always introduced by the giver of names when he wants to imitate what is *phusodes* (windy)." (*Cratylus*: 426 c--427) "Do you agree with me that the letter *rho* is expressive of rapidity, motion and hardness? And that *lambda* was expressive of smoothness, and softness, and the like?' (*Cratylus*: 434 c)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

symbolizes,[10] or an explanation of things via their names.[11] However, the value of the dialogue is not so much in its conclusions as in the fact that it raised certain questions so early and so authoritatively. Moreover, its arguments established a firm connection between intuitions of lexical fittingness and the phonetic contributions to those intuitions made by individual sounds.

Tracing the full history of thought on phonetic symbolism that stretches from Plato to Sapir is a vast, separate research task in itself, and its inclusion would unnecessarily complicate this study. What is worth noting, en passant, on our way to the 20th -century treatment of this topic are notable points of resemblance between Plato's description of speech sounds and two subsequent English-based descriptions, one put forward and the other quoted by the mid-19th century writer, Edwin Guest (1838: 12) in the chapter on Imitative Sounds in his *History of English rhythms*.[12] Here, we find an observation taken from Sir Francis Bacon's *Sylva sylvarum, or A naturall historie: in ten centuries* (1631), Bacon noted the ability of speech sounds ("letters") to mimic natural sounds; his interest is more onomatopoeic, perhaps not as abstract as Plato's intuitions about segments, but there are distinct resemblances: "There is found a similitude between the sound, that is made by inanimate bodies, or by animate bodies who have no voice articulate, and divers letters of articulate voices; and commonly men have given such names to those sounds as do allude unto the articulate letters; as trembling of water hath resemblance with the letter *l* ;

---

[10] "Words should as far as possible resemble things, but I fear that this dragging in of resemblance... is a shabby thing, which has to be supplemented by the mechanical aid of convention with a view to correctness." (*Cratylus*: 435 c)

[11] "The knowledge of things is not to be derived from names. No, they must be studied and investigated in themselves." (*Cratylus*: 439 b-c)

[12] For the reference to Guest, (and hence also to the observations of Bacon and Johnson) I am indebted to Professor William Stewart.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

quenching of hot metals with the letter *z* ; snarling of dogs with the letter *r* ; the noise of screech owls with the letter *sh*, voice of cats with the dipthong *eu*, voice of cuckoos with the dipthong *ou*, sounds of strings with the dipthong *ng*." Jumping to another century, Guest also cites Dr. Johnson (1709—1784), who remarks upon the imitative potential of speech sounds (in literature), though without giving phonetic details.[13]

Guest himself argued for including articulatory gesture, in contrast with Bacon's focus on the acoustic aspect of phonetic imitation: "If the mere sound of the words *hiss* and *bah* recall the cry of the animal, so may the muscular action, which the organs exert in pronouncing the words *struggle*, *wrestle*, call up the play of muscle and sinew usual in those encounters. *Wherever there is resemblance there may be association*"[14] (Guest 1838: 17, my emphasis). He also described the expressive effects provided by different groups of consonants, illustrating his points with examples from English poetry. Following examples of the effects of "the sounds represented by *z*, *zh*, *s*, *sh*" he remarked that "In several of these examples the sharp sound of the sibilant is strengthened by that of the narrow[15] vowels, long *e* and short *i*. These vowels are sometimes used with effect even by themselves... The broad vowel-sounds on the contrary, long *a*, *au*, long and short *o*, together with the broad dipthong *ou* are used to express deep and hollow sounds... The very expression a *hollow* sound shows how close is the association of a hollow space with depth and fullness of

---

[13] "Even Johnson, notwithstanding the ridicule he has thrown upon enquiries of this nature, has admitted that particular images may be "adumbrated by an exact and perceptible resemblance of sound." (Guest, 1838: 17)

[14] Guest's observation of articulatory detail and its mimetic potential should also be emphasized: "When we pronounce the letter l, the breath in escaping under the side teeth presses against the yielding tongue which may be considered as fixed at its root and tip. The tongue, like other flaccid bodies in similar circumstances, vibrates with a slow and uncertain trembling. This strongly resembles the motion of water." (Guest 1838: 12)

[15] Guest (1838: 6) differentiated between what he called "narrow" sounds (the high front vowels [iy], [ɪ] and [ey]) and "broad" ones (the back vowels [a], [ɔ] and [ow])

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sound. Hence the broad vowels are sometimes used to express mere breadth and concavity." (Guest 1838: 16).

In 1887, Albert Tolman proposed a "table of the English vowel colors," [16] (Tolman 1887: 327) in which vowels and dipthongs were ranked on a hierarchy, beginning with the mid high front vowel [ɪ] of "little" and ending with the mid low back rounded [ɔ] of "awe".[17] In fact, the germ of Tolman's vowel scale can be seen in the vowel distinctions made in the "Imitative Sounds" chapter of Guest's book, which Tolman knew and quoted from.[18] Tolman holds a special position in that whole line of observers of language who have noted the expressive properties of speech sounds. He stands out as having attempted to systematize his intuitions in the vowel scale, and as having claimed a scientific basis (the acoustic properties of vowels) for his system. And he also, both in his use of the term *vowel color* and in his use of feature clustering at the endpoints of the scale, connected the issue of "sound symbolism" (as he termed it) to that of synesthesia.

Tolman took Guest's view of vowel contrasts and expanded it: "The colors at the very bottom of this scale are peculiarly fitting to express solemnity, awe, horror and deep grief, also slowness of motion and extreme or oppressive greatness of size. The entire lower half of the scale, in general,

---

[16] Consonants were also treated in terms of their symbolisms in the article, but without the same ranking system as the vowels. (SeeTolman 1887: 332) Of interest here is the author's consciously synesthetic use of the word "color" to describe vowel quality, which is very much of his *fin-de-siècle* era.

[17] It should be noted that Sapir (1927) also chose [ɔ] as the largest of the three vowels described in his informal study. However, in his experiment (1929), [a] was found to be the largest vowel.

[18] "Dr. Guest ('Hist. English Rhythms') declares: 'Shakespeare seems to have affected the short vowels and particularly the short i when he had to describe any quickness of motion.' This is natural, 'Large bodies move slowly' and heavily, and their motion is best described by slow, heavy vowels; and quick movement by light vowels." (Tolman 1887: 330)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is adapted to express largeness, slowness, seriousness and gloom. The colors at the very top are especially fitted to express uncontrollable joy and delight, excessive gaiety, rapid movement, delicacy and physical littleness. The scale runs, then, from the large to the little, from horror to ecstatic delight, and from the solemn and awful to the simply trivial." (Tolman 1887: 327-8) In a later article treating the same material, Tolman also added the qualities of "brightness" to the top end of the scale and "darkness" to the bottom end. (Tolman 1904: 153) It is also in the later article that Tolman first used the term "sound symbolism," and here too that he clearly stated that if all speech sounds have symbolic potential, they are also complex and may lend themselves to more than one symbolism. [19]

The order in which Tolman arranged the vowels seems to have been prompted both by his own intuitions as a literary critic and by his reading of the phoneticians of his day. Ranking was based on what he termed "inherent pitch", by which he appears to have meant the variations in fundamental frequncy intrinsic to the vowels. He hypothesized that inherent F_0 variation would trigger consistent emotional associations: hence the possibility of predicting where vowels would range on a scale.[20] Although claiming that "the scale follows a definite physiological order" (Tolman 1887: 328), with neither the front nor the back vowels did he keep to a simple high-to-

---

[19] "The most subtle form of expression through the sounds of language... sometimes has been termed tone-color; it will often be called here sound-analogy or sound symbolism." (Tolman 1904: 150-1) "Every English sound has some special expressive force. *Also, since a sound may have many striking characteristics, it may have more than one natural expression.*" (Tolman 1904: 165, my emphasis)

[20] "Phoneticians recognize each vowel color as having certain inherent or natural pitch as compared with the others. My scale follows substantially the order of inherent pitch given by Sweet and others beginning at the bottom with the sounds of lowest natural pitch. What is more natural than that these individual vowel sounds should in time, be felt to be, *according to their natural pitch*, the best sound-representative of these (i.e. gloomy versus joyful, bright versus dark and so on) feelings and ideas? What is more certain than this result unless some strong counteracting tendency be present? It will be seen that I am considering inherent pitch as, practically, a part

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

low or low-to-high arrangement. Tolman's tone-color scale is presented below. I am assuming its author had an r-less pronunciation of "fair," and rendered the nucleus of "what" as the low back rounded vowel of British *hot*, *rock*, and so on.[21]

TONE-COLOR

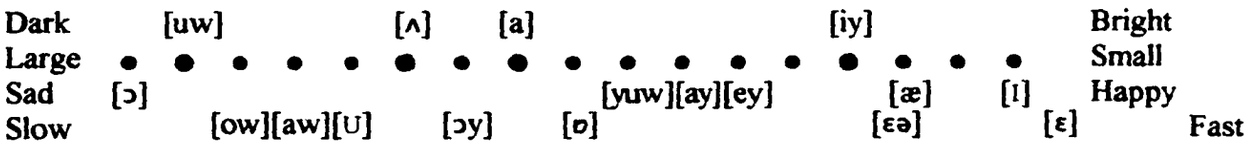


Figure 2.1, Tolman's tone-color scale. Tolman's (1904: 52) order for 17 American English vowels on a composite of attributes; spacing is arbitrary. The four stimulus vowels of this experiment are enlarged and presented above the scale. Below the scale on the first tier are vowels also used on Newman's (1933) brilliance and size scales (see below, p 36); the second tier presents only those vowels used by Tolman.

A notable aspect of the tone-color scale is that it assumed a *clustering* of semantic attributes at either end. In other words, the qualities of brightness, smallness, joy, rapid movement and so forth were in some sense seen as closely enough related that the same vowel might be chosen to symbolize them; the same could be said of the opposite qualities of darkness, largeness, sadness, slowness and so forth. What is suggested here is the idea of emotional-sensory equivalence, very close to the concept of synesthesia, which is primarily related to eqivalence among two or more sensory modalities (although of course the emotions too are connected to sense-modes). The

---

of tone-color." (Tolman 1887: 328)

[21] As evidenced by their separate status on his scale, Tolman clearly percieved the vowels of *what* and *but* as different. He was an educated late-Victorian American English speaker and a Shakesperian scholar, so there would be social and historical reasons for expecting an R.P vowel in *what*. I have been advised in this and on the r-lessness of "fair" by Professor William Stewart, in a personal communication.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

canonical example is *l'audition colorée* where each of the vowels has, for the listener, its own specific corresponding color.[22]

There are two ways that synesthesia can be seen in relation to phonetic symbolism. We can look at it firstly from the aspect of the signified, as in the case of Tolman's vowel scale, where different referents—e.g., the attributes brightness, rapidity, joy, smallness and so on—are considered to be related to each other. Tolman's cluster suggests not only a straightforward correspondance between the senses of sight and that of touch (kinesthetic-proprioperceptive) but also a connection to whatever personal sense experience is evoked by the emotion of joy in the listener. Then, secondly, when a signifier (e.g., a high front vowel) is found to "mean" these qualities, a further equivalence is established between an auditory object (i.e. one which evokes yet another sense) and the senses of sight, touch, emotion of joy and so on.

However, it would be oversimplifying to think of a speech sound as merely an auditory object since several senses are employed in its transmission and reception. A segment is produced via at least two[23] modalities: the *kinesthetic-proprioperceptive* (static and dynamic articulatory qualities:

---

[22] Synesthetic examples from mid to late nineteenth century decadent writers abound; they are also found in the work of modernists, such as Joyce, Sitwell and Cummings. Nabokov's is a well-known contemporary case. Famous earlier examples of synesthesia are Baudelaire's "Correspondances" and Rimbaud's "Sonnet of the Vowels." Yet another is Poe's poem "The Bells", of which Tolman says, "Every part is both imitative and symbolic... Poe wrote this poem consciously upon the distinct theory that those language sounds which best imitate some natural sound, here a bell of a particular tone, *are intrinsically adapted to express at the same time certain allied emotions...* (Tolman 1904: 168-9, my emphasis). Synesthesia is perhaps most strikingly exemplified in Huysmans' *Against Nature* (1959), originally published in 1884, three years before Tolman's first article. This strange and influential novel foregrounded the synesthetic experience of its jaded hero; for illustrative instances, see particularly the passages on the "mouth organ," uniting liqueurs and the sounds of the orchestra, (Huysmans 1959: 58) and the section on the blending of perfumes as stimulants of auditory and visual sensation (Huysmans 1959: 122-8).

[23] It could be argued that there is a third, *internalized visual* aspect to the production of a speech sound, in that the speaker may have some unconscious pictorialization of how his or her own lip and jaw movements appear to the listener.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

touch and movement in the oral tract) and the *aura*l (acoustic feedback through the bones of the head and through air vibration). The same sound is recieved through three modalities: the *visual* (lip and jaw movements of the speaker), the *aural* and what could be called the *imitatively kinesthetic-proprioperceptive*, in which the listener reduplicates in her or his own mouth the gesture by which the speaker has produced the perceived sound.

Clearly, this is a rich array of sensory experience; it offers a number of different features on which to base any symbolism. Since a speech sound of its nature requires the connecting of two or three different sensory aspects to form its matrix, it would seem a particularly fertile place to generate cross-modal symbolisms. (It also follows that phonetically-based synesthesias—like that of *l'audition colorée*—should not be considered the products of simple matchings between the auditory and visual senses, given that a vowel exists in several sense dimensions.)

The study of synesthesia is in itself a vast topic. It is mentioned here only insofar as it has fed directly into those lines of thought which eventually generated Sapir's 1929 experiment. Interest in it seems to have run chronologically alongside interest in phonetic symbolism: it was growing at the same time as Darwin's (1872) ideas about the gestural communication of emotions were beginning to be applied to language, and that writers like Tolman were combining literary and phonetic intuitions about speech sounds. Moreover, experiments concerning synesthesia peaked somewhere between 1927 and 1940, around the time of the first great wave of experiments in phonetic symbolism.[24]

---

[24] See Marks (1978: 86—87) for a discussion of the role of synesthesia in the intellectual history of the 19th and 20th centuries.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Tolman represents that part of the enquiry into phonetic symbolism which bypasses the fittingness debate to focus on phonetic imitation, and he concentrates very much on the acoustic aspect of speech sounds. But thinkers about language were to be directed also towards the proprioceptive-kinetic aspect by the publication of an important book, contemporaneous with Tolman's first article: *The expression of the emotions in man and the animals* (Darwin 1872).[25] This was a small part of Darwin's whole evolutionary theory, focusing on gestural meaning; it was to have an enormous influence on ideas about language and about the universal use of movement symbolism in communication.

In this work, Darwin proposed "three Principles, which appear...to account for most of the expressions and gestures involuntarily used by man and the lower animals under the influence of various emotions and sensations." (Darwin 1872: 27) The first, which relates directly to the concept of phonetic symbolism, is "*The principle of serviceable associated Habits.*" (Darwin 1872: 27) This puts forward the idea that the primary strategy for communicating feelings can be found in the metaphorically-extended use of gestures which would otherwise be employed quite literally. Moreover, even if certain highly visible expressive movements are inhibited, they still leave their traces in the actions of muscles less accessible to the inhibiting will—communicative use is still made of the partial survivals of those repressed gestures.[26] It is in the working out of this first principle that we find a very plausible argument for a shared basis of iconicity in human

---

[25] I am much obliged to Professor William Stewart for mentioning to me this clearly seminal book, which is by no means a standard item in bibliographies on phonetic symbolism.

[26] For a description of the process, see Darwin (1872: 28).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and animal communication. Phonetic symbolism represents one linguistic outgrowth of this general iconic tendency.

Many illuminating instances of the first principle are given, drawn from the animal kingdom,[27] but there are also many human examples: "A vulgar man often scratches his head when perplexed in mind; I believe that he acts thus from habit, as if he experienced a slightly uncomfortable bodily sensation, namely, the itching of his head, to which he is particularly liable, and which he thus relieves. Another man rubs his eyes when perplexed, or gives a little cough when embarrassed, acting in either case as if he felt a slightly uncomfortable sensation in his eyes or windpipe." (Darwin 1872: 32) "Spitting seems an almost universal sign of contempt or disgust; and spitting obviously represents the rejection of anything offensive from the mouth." (Darwin 1872: 260) The survival of partial gestures due to the effect of inhibition (resulting in movements that still carry expressive force) is also graphically exemplified: "...As soon as some melancholy thought passes through the brain, there appears a just perceptible drawing down of the corners of the mouth, or a slight raising up of the inner ends of the eyebrows, or both movements combined, and immediately afterwards a slight suffusion of tears. A thrill of nerve-force is transmitted along several habitual channels, *and produces an effect on any point where the will has not acquired through long habit much power of interference*. The above actions may be considered as

---

[27] For example, "Kittens, puppies and young pigs... alternately push with their forefeet against the mammary glands of their mothers, to excite a freer secretion of milk, or to make it flow. Now it is very common with young cats and not at all rare with old cats.... when comfortably lying on a warm shawl or other soft substance, to pound it quietly and alternately with their forefeet; their toes being spread out and claws slightly protruded, precisely as when sucking their mother. That it is the *same movement* is clearly shown by their often at the same time taking a bit of the shawl into their mouths and sucking it; generally closing their eyes and purring from delight. This curious movement is commonly excited only in association with the sensation of a warm soft surface; but I have seen an old cat, when pleased by having its back scratched, pounding the air with its feet in the same manner; *so that this action has become almost the expression of a pleasureable sensation*" (Darwin 1872: 46-7, my emphasis).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

rudimental vestiges of the sceaming-fits, which are so frequent and prolonged during infancy." (Darwin 1872: 195, my emphasis.)

The body of Darwin's book was given over to observations of the operation of the three principles. Since it is based on communications deriving from specific body-character reactions and reflexes (such as blushing in humans), Darwin's third principle is not relevant to the discussion of phonetic symbolism.[28] However, his second "principle of Antithesis", is of great interest, since it builds by polarity on the first: "When a directly opposite state of mind is induced, there is a strong and involuntary tendency to the performance of movements of a directly opposite nature, though these are of no use; and such movements are in some cases highly expressive." (Darwin 1872: 28) Thus, further use can be made of a gesture once it has developed a stable meaning. First comes the metaphoric extension of what was once a literal response to some stimulus (such as human spitting or head-scratching, or feline claw-flexing). Then when its significance is agreed upon, the gesture becomes available for *opposition*, i.e. for a movement which is obviously the contrary to indicate a contrary feeling state.[29] Here Darwin seems to have

---

[28] The third principle is: "the principle of actions due to the constitution of the Nervous System, independently of the first of the Will, and independently to a certain extent of Habit." (Darwin 1872: 29) "The most striking case... of the direct influence of the nervous system... on the body, is the loss of colour in the hair, which has occasionally been observed after extreme terror or grief.... Another good case is that of the trembling of the muscles, which is common to man and to many, or most, of the lower animals. Trembling is of no service, often of much disservice, and cannot have been at first acquired through the will and then rendered habitual in association with any emotion.... Trembling is excited in different individuals, in very different degrees, and by the most diversified causes..." (Darwin 1872: 67) Meanings based on gestures derived from this principle would be doubly unsuitable for linguistic use. Firstly, they are not necessarily stable (as in the trembling case, cited above). Secondly, out of all the three gesture types, this third group are dependent on the body in the most literal, untranslatable, one-for-one way. Such movements do not carry the possibilities of synecdoche, where one part of some complex expressive gesture (most of which may have been repressed) can come to stand for the whole.

[29] For instance, the shrug, apparently universally used to mean detachment or impotence, arose originally in opposition to certain combat movements: holding the head erect, squaring the shoulders, expanding the chest, clenching the fists, rigidifying the muscles of the limbs and so on. As these fight movements came by extension

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

been describing essentially the same meaning strategy as de Saussure (1959: 122) did in the latter's observations about the functioning of oppositions in language. De Saussure however, made a wider claim for his concept, since he asserted that *all* linguistic forms derive their meanings from being contrasted with others in the relevant paradigms. Of course, since it is a linguistic, rather than a general gestural, system that de Saussure described, Darwin's first principle, that of "serviceable associated Habits," is not available to provide inherent meaning: Saussurean oppositional meanings come from the relation *not-x*, while *x* itself stays arbitrary.

It is easy to see only in the universalism in Darwin's book; it starts with broad prinicples of communication, with examples drawn both from human behavior and that of the "lower" animals. In the later sections, where there is a more detailed discussion about human strategies of emotional expression, Darwin took care to draw his examples from all nationalities and races. Moreover, his conclusions, though guarded, were definitely evolutionist, based on a universalist point of view: "We have seen that the study of the theory of expression confirms to a certain limited extent the conclusion that man is derived from some lower animal form, and supports the belief of the specific, or subspecific unity of the several races" (Darwin 1872: 365). Yet there is a big difference between describing the broad underlying strategies of communication (the three principles) and asserting that all their manifestations when expressing the same emotion will resemble each other. Darwin's focus on tracing generally shared behaviors never led him to claim that exceptions and variations do not exist. In the his writing on animals, he was as quick to point

---

(following the first principle) to signify not just literal, functional preparation for attack or defense but to show more generally a state of aroused indignation and a readiness to respond, so a contrary *non-functional* set of movements (shoulders held high, hands sometimes open or uplifted) became available to indicate helplessness or detachment (following the second principle).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

out differences as similarities in gestural strategies.[30] Naturally, in his analysis of human expression, Darwin focused on universal tendencies wherever possible. Yet his observations did not over-universalize at the expense of the truth: they still took cultural and racial differences into account.[31]

Still, however much Darwin himself was careful not to oversimplify, it has been easy when following his line of reasoning to fall into certain unquestioned assumptions of universality, all the more dangerous because they may never have been overtly stated. In his *Analytics of literature*, published two decades later, Sherman (1893) included a chapter on Tone Quality where the influence of Darwin, with its tendency to universalize, was very clear. Sherman stated: "The same principles (of expressive use of what he termed "phonetic qualities" or, following Tolman, "Tone Colors") are discernible everywhere, *and in fact prevail throughout the whole range of animal notes and cries...The best illustration of the universality as well as the potency of tone colors may be found outside and below the human scale*." (Sherman 1893: 24 and 27, my emphasis.) Like Tolman, Sherman was interested in phonetic imitation and not in lexical fittingness.[32] Also like Tolman, except his version was much simplified and included consonants, Sherman divided speech sounds into two opposing groups: "heavy" vowels and "hard" consonants versus "light" vowels and "liquid" consonants (Sherman 1893: 22). However, his focus was not merely acoustic—there was a Darwinian emphasis on gesture: "Clenched hands, set or grating teeth,

---

[30] For illustrative examples, see Darwin (1872: 127 and 136).

[31] For examples of both universal and culture-specific behaviors, see Darwin (1872: 211, 245 and 294).

[32] "The mind often gets an experience out of the phonetic elements of which words are made, independently of what the words as such may signify." (Sherman 1893: 21)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

involuntary rigor of muscles, tell their own story, even to the slowest imagination. The like is true also of vocal sounds."(Sherman 1893: 23)

A few years after the appearance of Tolman's (1904) second essay on the vowel scale, early experimental investigations on phonetic symbolism were beginning. Roblee and Washburn (1912) at the psychological laboratory of Vassar College investigated the "agreeable and disagreeable character of the sounds used in speech" (Roblee and Washburn 1912: 579) The focus of their experiment could be best described as fairly abstract, since pleasingness is only indirectly connectable to mimicked qualities. Stimuli were VC nonsense syllables; the experiment was specifically designed to evade lexical associations. Stimulus presentation was oral, with care taken to keep "the same pitch and intensity of voice throughout." (Roblee and Washburn 1912: 579) In all, 16 consonants and 12 vowels were tested. As there were only a small number of subjects (15 Vassar College students, presumably English speakers), any results are only of anecdotal interest, although of possible significance for this study was the fact that [ʌ] was judged the most unpleasant vowel and [a] (together with [ɛ]) the most pleasant (Roblee and Washburn 1912: 580).[33] The ranking of the evaluated vowels, as ranged along a pleasant-unpleasant continuum, did not follow any notable acoustic or articulatory order.

Robbins, Smith and Washburn (1915) carried out a further study, again with orally-presented, non-lexical VC syllables, but this time using 14 vowels and 15 consonants; data was gathered from 65 mostly "young women college students." (Robbbins, Smith and Washburn 1915: 293)

[33] See Chapters 7-10, in which findings are reported from the English and Spanish groups that [ʌ] was the darkest/most passive stimulus vowel and that the Korean and Japanese groups judged [a] to be the brightest and most active.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The reasons for judging vowels as pleasant or unpleasant were the focus of this second investigation. "Associative suggestions which the sounds have in spite of their meaningless character" (Robbins, Smith and Washburn 1915: 293) and articulatory ease of pronunciation[34] were hypothesized as probable causes of evaluations. The experimenters found that extreme or very strong judgments were "more apt to be due to associations" (Robbins, Smith and Washburn 1915: 293) than moderate ones—but only for consonants. Associations were mostly lexical or onomatopoeic: "an object making a similar sound" (Robbins, Smith and Washburn 1915: 293). Conclusions were that vowels "do not derive their pleasantnes or unpleasantness from association with meanings or with kinesthesis."(Robbins, Smith and Washburn 1915: 295).

What is unsatisfactory in the design of both of the above experiments are the over-various combinations of stimulus vowels and consonants—multisegmental stimuli—which lack the control of phonetic variables that would have been given by the use of minimal pairs,[35] so that judgments could not be reliably compared. The observation of a contemporaneous writer on language and literature, Logan Pearsall Smith, is uncomfortably apt here: "This symbolism of sounds, the suggestive power of *various combinations of vowels and consonants*, has never been very carefully studied..." (Smith 1912: 52, my emphasis).

---

[34] The experimenters found that this second cause did not apply so much to vowels: "There is more of the purely auditory character and less of the kinaesthetic about a vowel than about a consonant." (Robbins, Smith and Washburn 1915: 294) But in the above statement they are assuming, perhaps wrongly, that the articulatory vowel features (e.g., lip gesture or tongue height) are not also kinesthetically significant qualities, albeit less obviously dynamic than the articulatory features of consonants.

[35] For example, as in the hypothetical experimental task of choosing the more pleasant syllable out of *ig* vs *id*, or *id* vs *ip*, or the most unpleasant out of *ig* vs *ag* ,or *ag* vs *og*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In his book, Smith restated Guest's and Tolman's conclusions that "Long vowels suggest a slower movement than the shorter vowels...vowels which we pronounce by opening the mouth convey the idea of more massive objects; while those which are formed by nearly closing the lips suggest more slight movements or more slender objects. Thus *dong* is deeper in sound than *ding*, *clank* than *clink*, and *chip* is a slighter action than that described by *chop*.[36] More subtle are the suggestions provided by consonants."[37] (Smith 1912: 52) In his emphasis on the symbolic contribution of gesture, Smith shows the influence of Darwin : "This suggestive power is due partly to direct imitation of sounds, but more to the movement of the vocal organs, and their analogy with the movements we wish to describe." (Smith 1912: 52-3) Focusing on the lexicon, he noted that there was a give-and-take between the salient phonetic symbolism of certain words and subsequent paradigm formation.[38] Of the coining of new words, he felt that *fittingness* was definitely a strong creative influence: "They (i.e., new words) seem to be creations of the most vital faculty in language, the sense of an inherent and natural fitness of the name with the thing." (Smith 1912: 54)

The fittingness of words, more specifically of "unknown proper names", was to become the subject of experiments by English (1916) and Alspach (1917). Their studies had been inspired by an article on this topic (Kollarits 1914) and its response (Claparède 1914) in the Swiss journal *Psychologie quotidienne*. Kollarits had examined associations to unknown proper names,

---

[36] Here real words are used, but they are essentially minimal-pair frames for the vowels.

[37] Thus, Smith comes to the *opposite* conclusion from Robbins et al.

[38] "A common word established in the language describes a sond or action, and its sound comes to be connected with the thing that it describes. Other words are formed on its model, and finally the expressive power of the sound, suggesting as it does, so many other words of similar meaning, becomes a part of the unconscious heritage of those who use the same form of speech." (Smith 1912: 53) For a contemporary similar viewpoint, see also

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

concluding that we supply them from specific memories. Thus the name of an unknown person might suggest the somewhat similar name of someone known to have a certain profession, be of a certain nationality and so on. "Visual images are an aid to thought" he claimed (Kollarits 1914: 247), but gave no reference to any type of phonetic symbolism as a possible stimulus to association. Claparède, responding in the same volume of the journal, corrected him: "There are some other factors (causative of associations) that should be mentioned... *the physiognomy of the proper name* assuredly plays a large part" (Claparède 1914: 301) He then gave a pair of hypothetical contrasting names and claimed that associative responses to these and other names from literature would be generally the same, i.e., not conditioned by individual preferences and associations.[39] The rest of the article explored the author's personal types of association to words and made no especial general predictions.

Claparède's article was introspective and non-experimental, but it raised issues that encouraged others to investigate. English (1916) ran three tests on word-associations with unknown proper names as stimuli, presented orally. Only a small number of subjects (7, 8 and 11) were used, and conclusions were that "the physiognomy of a proper name is at best but one—and that not an important—factor out of the many which determine what personality shall be attached to the name... The psychological response to proper names is extremely variable." (English 1916: 433-4)

Bloomfield (1895).

[39] "The sound of the name has an affective tone which co-operates in the elaboration of its mental representation. Other things equal, names consisting of heavy or repeated syllables call forth images of fat, heavy-set, bloated or slightly ridiculous individuals; a short and sonorous name, on the other hand, suggests a slender and active person etc. M. Patapoufard would evidently be a very different type from M. Flic.... It is not without intention that Daudet created the name of Tartarin, Dickens that of Pickwick, Flaubert those of Bouvard and Pécuchet... These names produce a similar effect upon all readers." (Claparède 1914: 302).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

One subject alone in English's third test "consistently invented names that by sound were intended to fit the individual." Since this subject had shown a notable tendency to be responsive and inventive with sounds, fifteen months later he was made the sole focus of an experiment by Alspach (1917). The same 50 artificial proper names that had been used in English's third test were again presented auditorily to the subject. He was then to describe a hypothetical person fittingly named by each name; he was also required to give an introspective report on the processes by which he had arived at the name's imaginary referent. The subject reported that 35% of his assignment of names was due to "the mere sound of the word." (Alspach 1917: 438) Associations were also reported due to kinesthesis and to visual processes.[40] From the results, Alspach extrapolated two mixed[41] sound-groups, reminiscent of the attribute-cluster endpoints of Tolman's vowel scale: one *big, smooth, slow, broad, generous* and *soft*, and the other *snappy, active, clean-cut, definite, thin, sharp* and *hard*. Unfortunately, the distribution of the sounds in the groups was not particularly consistent, although there was a tendency for back and rounded vowels to be in the first group,[42] while [I] was the one consistent vowel in the second. The subject's introspections are interesting; however, since he was the only one tested, the experiment has only anecdotal significance. Moreover, using multisegmental stimuli brought its own confusions—since each stimulus word was a complex entity, reading the subject's introspective reports, it is impossible to analyze his responses in simple phonetic terms.

---

[40] For example, "*Witaw*: a visual image of a dog; tended to set teeth or snarl.... *Lisrix*: slips easily off your tongue; no effort.." (Alspach 1917: 442).

[41] Confusingly, these contained full names, consonant clusters, syllables and individual sounds.

[42] For support of this grouping, see the dark-slow-sad etc. end of Tolman's vowel scale in Figure 2.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

It was also in this period[43] that Jespersen (1922) included an entire chapter on "Sound Symbolism" (i.e. what is termed here phonetic symbolism) in his book, *Language*. Jespersen's position was balanced and moderate, neither totally opting for arbitrariness nor denying that phonetic symbolism was present and at work in language (Jespersen 1922: 397). Phonetic symbolism was classified into types, with numerous, mostly Indo-European, lexical instances for each category, starting with the most concrete, onomatopoeia ("Direct Imitation") and going on to more abstract applications ("Movement", "Things and Appearances", "States of Mind" and "Size and Distance"). There was also a section on the expressive "Length and Strength of Words."

Jespersen remarked that there was a "natural association between high tones (sounds especially marked by very rapid vibrations of acoustic energy) and light, and inversely between low tones and darkness" (Jespersen 1922: 400) He observed that the use of the vowel [iy] was associated especially with smallnes and with proximal forms, while [uw] and [a] were associated with distal ones (Jespersen 1922: 401—402). Still holding to a moderate view, examples were given of the complex interplay between phonetic symbolism and the lexicon.[44] He concluded that phonetic symbolism was an increasingly active force in language; it "makes some words more fit to survive and gives them considerable help in their struggle for existence." (Jespersen 1922: 408). "These forces are vital in language," resulting in "a greater number of easy and adequate expressions... in

[43] See also the later Jespersen (1933) article On the symbolic value of the vowel [iy].

[44] Jespersen noted that lexical movement may occur in both directions: words may either develop phonetic symbolism, or can lose their symbolic expressiveness due to semantic or phonological change.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which sound and sense are united in a marriage-union closer than was ever known to our remote ancestors" (Jespersen 1922: 409).

What is of note in Jespersen's writing is his refusal to follow unquestioningly de Saussure's line on arbitrariness. Having quoted de Saussure on the language-specific phonological and morphological constraints on onomatopoeia, he observes: "Modern linguistic science... is so preoccupied with etymology, with the origin of words, that it pays much more attention to what words have come from than to what they have come to be. If a word has not always been suggestive on account of its sound, then its actual suggestiveness is left out of account and may even be declared to be merely fanciful." (Jespersen 1922: 410). There is an attempt here to integrate de Saussure's undeniable insights into arbitrariness[45] with the sense that phonetic symbolism is an important aspect of linguistic drift,[46], more specifically of lexical formation and preservation.

Five years later, Sapir (1927) was to publish an article on phonetic symbolism that continued Jespersen's line of thinking—not to negate de Saussure's discoveries, but to balance them with observations drawn from the phonosymbolic approach to segmental and lexical meaning which has been surveyed in the preceding pages. For Sapir, a purely Saussurean view of the signifier-signified connection had come to seem overly one-dimensional. He argued that in language, underlying the system of arbitrary symbolization there lay another, highly determined "primary

[45] "No language utilizes sound symbolism to its full extent, but contains numerous words that are indifferent to or may even jar with symbolism... Words that have been symbolically expressive may cease to be so in consequence of historical development, either phonetic or semantic or both." (Jespersen 1922: 408) This last is exactly de Saussure's point (see again de Saussure, 1959: 69).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

gesture symbolism" (Sapir 1927: 430) in which vocal tract movements iconically represented the meanings of the speaker. The top-layer, arbitrary connection between linguistic forms and referents he termed the "referential" or "denotive" aspect of language (Sapir 1927: 425); the underlying, iconic system of relating form to meaning he called "expressive." (Sapir 1927: 426) "Language... has embodied two distinct strata of symbolic expression, a primary gesture symbolism and a secondary referential one which has largely swamped the former... all normal speech involves a conflict between, or intertwining of, two distinct strata of expression or types of mechanism..." (Sapir 1927: 430-1)

Using the example of a repressed angry impulse to strike someone, he explained, "I may inhibit the reaction without entirely destroying its form. There may be enough of the gesture left to make it obvious to the bystanders that I had the impulse to strike. This abbreviated or rudimentary gesture will then be felt as in some sort a symbol of my attitude.... Suppose further, that this gesture becomes accepted by society at large as a pantomimic symbol of the striking act, then we have what amounts to... an unemotional, denotive symbol of the notion of striking. It would then be possible for human beings to use such a gesture for purposes of reference only... *What if speech were, at last analysis, a highly conventionalized system of just such gestures which by constant use had lost their original expressive content, having been rubbed down to the status of purely referential symbols?*" (Sapir 1927: 426—427, my emphasis).

This is exactly Darwin's point, and Sapir then developed his argument along classic Darwinian lines, examining the influence of repression and of the resulting synecdochic relation of surviving

---

[46] For the concept of language drift, see Sapir (1921: 147-70).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

peripheral or partial gesture to message.[47] From this point in his reasoning, it was a very small step to the question of how a specific symbolism, for example that of size, might operate in language: "One's natural impulse in describing a large object is to move the arms in a more or less circular fashion suggesting the notion of bigness...The speech apparatus would help along in the expressive symbolism by parting the lips, dropping the tongue as low as possible, rounding the lips, sinking the pitch of the voice, and in other ways...If anyone is inclined to doubt the reality of such involuntary symbolisms in speech, let him try the following experiment, which I myself have tried with practically one hundred per cent success..." (Sapir 1927: 429)

Then followed an informal account of a technique to be published shortly afterwards as part the methodology of his pioneering phonetic symbolism study (Sapir 1929). Here, three nonsense words all meaning "table"—syllables with contrasting vowels, "lee", "la" (rhyming with "pa") and "law"—were to be judged by subjects in terms of their relative size. "I think it will be found that the normal English-speaking person, or French-speaking person, for that matter, will think of "lee" as symbolizing the small table, "law" the big table, and "la' the middle-sized table *par*

---

[47] "It is true to say that the whole body participates in the response in some form or other... we have reason to believe that the particular part of the body that carries out a given act is always accompanied by other segments of the organism, which lend their consent, as it were, in symbolic form... If the central expressive organ or system of organs is inhibited, the energy of the impulse may concentrate in some other part of the organism, so that the secondary expressive symbol may become *the* symbol *par excellence*. The more highly evolved we become as socialized individuals whose business it is to inhibit many of our expressive impulses, the more likely it is that these impulses find lodgement in parts of the organism that were not primarily designed, so to speak, for such expressions... If I cannot box a man on the ear, I can at least... quietly stick out my tongue at him when he is not looking... The tongue darts forward as a functionally ineffective but symbolically relieving substitute for the striking act itself... We may surmise that in the remote past speech was neither more nor less than a series of auditory gestures, directly symbolizing various types of adjusstment to the environment... substitutive symbols for other activities or other expressions which it was not possible or convenient to effect at the time" (Sapir: 1927: 428).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

excellence... Even at this late day we have not lost the feeling for the gesture significance of sounds and combinations of sounds." (Sapir 1927: 429) [48]

By claiming the existence of a linguistic substrate that was non-arbitrary and expressive, by emphasizing the role of repressed movements, and even by his use of the term "expressive", Sapir was clearly extending Darwin's line of argument on behavior in *The expression of the emotions* to language. And, significantly, his assumption seems to have been that this expressive substrate was universal: "We must learn to see languages as possessed of certain essential characteristics apart from those of particular languages that we may happen to be familiar with and *as rooted in some general soil of behavior that gives birth to other than strictly linguistic forms of expression*."(Sapir 1927: 421, my emphasis.) The underlying reasoning seems to be as follows: whatever general expressive gestural strategies we share with the animals and with those humans of other races and cultures must somehow generate the same behavior on the most concrete level.

In the same year that Sapir published his article, two others appeared in the journal *Psyche* as if in support of his views on phonetic symbolism. These were by Paget (1927), and by von Hornbostel (1925, 1927), translated from German. Paget, although making no overt reference to evolutionary theory in this article, was clearly much influenced by Darwin's thought[49] in his hypothesis about the gestural basis of language: "When we talk, we are symbolizing our thoughts... by making a

---

[48] Even though here Sapir was using syllables (i.e. multisegmental stimuli), he was also using *minimal pairs*, so that any symbolic effect due to the initial liquid was neutralized.

[49] For confirmation of this, see Paget's *Human speech* (1930). In support of his 'gesture theory of speech', the author quotes from the very section of Darwin's *Expression of the emotions* which introduces the "Principle of serviceable associated Habits". (Paget 1930: 133).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

kind of sign-language, and... when we listen to another person talking we are reading his sign language 'by ear'.... I believe... that man started to speak by imitating with his tongue and other vocal organs the gestures which he had previously made with his hands, or rather with his hands and face together." (Paget 1927: 37)

Like Sapir, Paget chose as his first example of linguistic gesture to describe those oral strategies used in communicating size: "Let us see how such a plan would work out in practice. His (i.e. primitive man's) old, original gesture for 'little' would be to represent something small with his hands. His corresponding speech gesture would therefore be to make a small mouth, bringing his tongue forward near his lips and partly closing his lips so as to make a small cavity. His hand gesture for 'big' would be one which represented something big; so his speech gesture would be to make a big mouth, with large opening of the jaws, and tongue far back... The little mouth... makes a 'word' like *ee-ee* or *ii-ii* (as in French) according to the extent to which we close our lips. The big mouth makes a word like *aw-aw* or *oh-oh* in the same way."(Paget 1927: 37) Again, as with Sapir, an obvious universalism was implied: the underlying assumption was that for each possible parameter of oral position (lip gesture, tongue height and so on) there could be only one choice which would correctly convey a message of largeness or smallness.

Although predating it by fifty years, Darwin's concept of "serviceable associated Habits" stands directly behind Sapir's and Paget's idea of language as symbolizing gesture. However, interest in synesthesia was also mounting at this time, and another article in *Psyche* "The unity of the senses" (von Hornbostel, 1927) makes important connections with the idea of sensory equivalences or correspondances, and with the non-gestural, acoustic aspect of speech sounds. Its author was an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

acoustician, but his style here was more literary or philosophical than technical: "It matters little through which sense I realize that in the dark I have blundered into a pig-sty... Everyone knows what 'brightness' of sound means—not something corresponding to light, but the same thing....Anyone can find on the piano that tone which sounds as bright as lilac smells. (Generally he thinks the task nonsense at first, but if he can be persuaded to deal with such nonsense at all, it goes very well.) ...So there is a sensuous which is not limited to one single sense. Indeed, looking more closely, the apparent exception becomes the rule, and one must search in order to find the private property of any one sense."(von Hornbostel 1927: 83—84)

Like Sapir, Von Hornbostel viewed both lexical fittingness and phonetic imitation as survivals of an earlier type of more unified, primitive world-view.[50] He was yet another writer on phonetic symbolism who saw speech sounds as classifiable into two groups with opposing clusters of qualities: "The sense of hollow sounds (like m, mb) is 'dark, dull, bitter, blunt, heavy, dense, thick, big, full, round, swelling, deep, tired', and much more, but all this in one. And now think of the opposites, 'bright, sharp, light, blank' and so forth—how strange it would be for these to say that they were 'umb'! Nobody would believe them. Now speech is not the setting of words one after another, but is a happening in sound. *Even single isolated sounds have a sense*—the example was only meant to show that they do have one—*but it is less definite than in the course of speech, and often essentially different.* It is only the structure of this course, the melody, which

[50] Thus, von Hornbostel writes of primitive man., "In his perception are desire and fear; his thoughts stand before him, and behave as any of his kind behaves. He does not put soul into things, because soul has not yet been taken out of them. He does not pin names on to things, does not ask, "What are you called?", but "Who are you?" *because everything still is what it is called.* And as it is, so it looks, so it sounds, so it feels, so it does to him for good or evil, and so he does to it in turn.*This 'so' is heard in speech. The sound paints-- and paints more than mere sound.*" (von Hornbostel 1927: 86—7, my emphasis). There is a striking likeness here to his contemporary, Piaget's, (1929) descriptions of the beginnings of thought in children, most especially of the stage of Nominal Realism.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

transposes the living reality into the sphere of acoustics, at the same time leaving its full sense intact" (von Hornbostel 1927: 87, my emphasis).

At this point, a study was needed that would methodically investigate what Socrates had first pointed out in his dissection of "primary forms" into sounds: the *phonetic basis* of fittingness. And so, Sapir himself answered the challenge he himself had thrown out two years previously: "Students of speech psychology could hardly do better than work out experiments intended to test to what extent imaginary speech forms can be constructed that are capable of intuitive interpretation in a primary symbolic sense." (Sapir 1927: 403) The advance that Sapir's (1929) experiment represented over all of the previous experiments mentioned above was in his use of minimal pair artficial stimuli (nonsense words), thus excluding lexical influences and specifically focusing on attributions through use of linguistic oppositions. He saw that his task was not simply a question of finding a quality which might be unconsciously assigned to a speech sound, but setting up both sound (stimulus) and quality (attribute to be assigned) in terms of a series of oppositions—a very Saussurean view for someone challenging de Saussure's central tenet of arbitrariness.[51]

Sapir decided to test "the meaning contrast 'large': 'small' as the most likely chance of arriving at tangible results." (Sapir 1929: 226) There were three series of tests, in which both vowels and consonants were examined. In the first test, size judgments were elicited on two stimulus sets, the

---

[51] See N. Ciccotosto (1991: 147): "Since whatever innate tendencies exist for sound symbolism might be expressed in sound contrasts rather than merely one phoneme, Sapir devised a number of artificial lexicons to represent the range of high-low and front-back vowels for English."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

stimuli being pairs of artificial three-segment words. The basic elicitation technique—i.e., comparing the word-pair members to see which seemed larger—had already been described in Sapir (1927). One set of 30 tokens contained only English speech sounds, while another 30-token set also contained some non-English speech sounds. Presentation was presumably oral, but was not specifically described in the article. The example of stimulus words given were *mal* and *mil* (Sapir 1929: 227), a minimal pair. Thus Sapir was using the standard technique for discovering the phonological inventory of an unknown language by continuing to use word-like sequences (i.e. strings that felt somewhat natural) yet still being able to focus on one segment at a time. The importance of this application of the minimal pair technique cannot be overemphasized.[52]

Results from the first tests indicated that there were indeed consistent size attributions to both vowels and consonants, and that they were irrespective of "whether the phonetic contrast was contained in a phonetically 'possible' or a phonetically 'impossible' context." (Sapir 1929: 228) However, these types of tests were replaced because of "leading to a too ready systematization of responses on the part of the subject", with a "more efficient" test series (Sapir 1929: 228), using 100 artificial word pairs in a completely randomized order. The second tests were given to 500 subjects, a large number of whom were school children.[53] Five groups of speech sounds were tested, both vowels and consonants. Presentation was oral, and instructions were that "if the

[52] Where researchers have departed from this model, since even the most minimal CVC stimulus will provide three matrixes of features (and, in addition, the extra possible significance of the consonants vis a vis their placement in the syllable), symbolic responses cannot not be reliably correlated with any one segment. Bentley and Varon, for example contrasted the syllables *fam* and *jif* for largeness, and *miv* and *vag* for softness (Bentley and Varon (1933: 85)). If they wished to examine the contrast [a] : [i], why complicate the issue by including nasality, voicing, frication, affrication, plosivity, labiality, velarity and coronality, plus pre- and post-vocalic positioning?

[53] Ages ranged from 11 to 18, with a majority being between 13 and 17. There were also 29 adults (8 teachers of the children tested, and 21 university students) and 7 Chinese (adults).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

response was indifferent, no check was to be entered into either... column." (Sapir 1929: 229) In other words, given that this was a forced-choice test, provision was made for subjects to find the notion of size quite irrelevant to both stimuli, and to opt out by not choosing either stimulus word as the larger or the smaller.

In his article, Sapir focused only on the [a] : [i] contrast[54], since in his words: "It would be quite impossible to report on all the details of the experiment in this place." (Sapir 1929: 230) He observed that "English-speaking society does, for some reason or other, feel that of these two vowels, *a* by and large is possessed of a greater magnitude symbolism than the contrasted vowel *i*. The same feeling seems to be illustrated by the small number of Chinese cases." (Sapir 1929: 231)[55] The age of the subjects did not seem to be a factor: "There seems little reason to believe that there is significant growth in the firmness of symbolic feeling after the age of 11." (Sapir 1929: 231)[56]

Sapir also observed that if the degrees of height of the four front vowels and the lowness and backness of the one back vowel in the group are considered as "steps", so that between the extremes [i] and [a] there are four steps, then the greater the contrast (i.e. the greater the number of steps) between two vowels, the greater the unanimity of judgments that the higher vowel was smaller, and that the lower vowel was larger. However, in this second set of results, an interesting anomaly occurred. The contrasted set [a] : [e] was judged with "some factor of hesitation" even

---

[54] 5 vowels were tested in this particular series: German [a], English [ae], English [E], French [e] and French [i].

[55] Since the Chinese numbers were tiny, this was clearly only the most provisional of conclusions for a non-Indo European speaking population.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

though it was a "three-step one." (Sapir 1929: 234); this vowel pair evoked surpirsingly low responses, more like those of a two-step or a one-step contrast. Sapir's explanation was that his English-speaking subjects had been conditioned by their native phonology to hear [e] as [ey], and to interpret it as "large" due to their perception of it as a long (or tense/dipthongized) vowel. He observed that this example illustrated "the importance of of the linguistic factor vs. the merely phonetic one", in other words the influence of "the *phonetic configuration* of English as such." (Sapir 1929: 234). This anomalous case in Sapir's experiment has, to my knowledge, not been commented on in any of the subsequent literature on phonetic symbolism. However, it suggests very strongly that the subjects' native phonology can set definite limits to the universality of responses to be expected, and that more than one phonetic factor may be responsible for producing phonosymbolic judgments.

This case apart, Sapir observed that "symbolic discriminations run encouragingly parallell to the objective ones based on phonetic considerations" (Sapir 1929: 233), and that "the reason for this unconscious symbolism may be acoustic or kinesthetic or a combination of both." (Sapir 1929: 235) It was also noted that "individuals differ a good deal in the matter of sensitiveness to the symbolic suggestiveness of special sound contrasts." (Sapir 1929: 237) Moreover, he found that what triggered the symbolizing reaction could be different aspects of the speech sound, depending on the individual; when discussing the experiment afterwards with his child subjects, "the impression was gained that the subjects differed somewhat in the psychological basis of the symbolism, some being apparently swayed entirely by the acoustic factor, others by the acoustic

---

[56] Newman, analyzing the same data came to slightly different conclusions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

factor only or mainly insofar as it was itself supported by the kinesthetic factor."[57] (Sapir 1929: 236)

Finally "a third experiment, intended to focus on individual idiosyncracies was carried out with a number of selected subjects." (Sapir 1929: 236) An artificial word was given and assigned an arbitrary meaning. Then a speech sound in the word was changed and the subject asked to report on the resulting meaning difference. There was much variation in ability, but some subjects were able to form what amounted to minimal paradigms of up to 50 words in a "constellated system" (Sapir 1929: 236) in which meanings varied due to phonetic change one segment at a time. Moreover, although size symbolism was not suggested by the experimenter, "a considerable number of the responses" accorded with "magnitude symbolisms independently obtained" (Sapir 1929: 238) from the earlier two tests. Sapir concluded that "The experiments here referred to give cumulative evidence for the belief that unsocialized symbolisms tend to work themselves out rather definitely" (Sapir 1929: 239), i.e., that sound evidence for the existence of phonetic symbolism had been discovered.

Four years later, Newman used some of Sapir's data and collected some more of his own in "an attempt to study the patterning of phonetic symbolisms on a non-linguistic plane." (Newman 1933: 53)[58] Minimal pair stimuli were again used. In the first experiment, Sapir's (1929) data from the second series of tests was divided by subject age into three groups: 9-13, 14-15 and 16 plus. The oldest group showed the most unanimity of responses, and the youngest group the least; it

---

[57] This observation accords with the introspective reports of Alspach's single observer, if he is taken as a kind of Everysubject.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

displayed the same general patterning, but tended to "discriminate less sharply and less consistently...between the sounds." (Newman 1933: 59)[59] Apart from reversals with the pairs [a]/[u] (youngest group) and [o]/[ɔ] (middle group), the ranking from large to small was agreed as follows: [ɔ], [o], [u], [a], [æ], [ɛ], [e], [i]. There were some anomalies in the data, but Newman noted that they seemed to have their own regularity: "Discrepancies point not only to chance fluctuation but to constant shift," and although admitting "The reasons cannot be satisfactorily explained," he insisted that "We are evidently dealing with a field of latent symbolism in which the patterning is even more rigid and constant than the discrepancies would lead us to believe." (Newman 1933: 60-1)

Three factors "of a mechanical nature" were outlined as responsible for judgments: "(1) the receding positions of articulation made by the tongue within the mouth, (2) the decreasing frequencies of vocalic resonance as measured acoustically and (3) the increasing size of the oral cavity used in pronunciation." (Newman 1933: 61) It was noted that the "three factors were at variance in their arrangement of the back vowels" although "in unison" (Newman 1933: 61) for the front vowels. The first, kinesthetic, "front to back" and the second, acoustic, factors ranked the back vowels from small to large as follows: [a], [ɔ], [o], [u], while the third "small to large", "kinesthetic, possibly visual factor" (Newman 1933: 62) ranked them [a], [u], [o], [ɔ]. It was observed that "the symbolic magnitude pattern conforms to the first two series in placing the back vowels above the front vowels, but it compensates for this arrangement... by following the sequence of the third series." (Newman 1933: 62) This is an early indication that the different

---

[58] There is a certain confusion of terminology here: by "non-linguistic" Newman means *non-lexical*.

[59] This is a little different from Sapir's conclusions drawn from the same data.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

'factors' (or dimensions, or features) of speech sounds might at points be in conflict with each other or produce different rankings for different scales (i.e., qualities).

In a second experiment, two symbolisms were tested: large/small and bright/dark, examining both vowels and consonants. A large number of subjects were used.[60] In contrast with the Sapir's experiment, "elements were restricted to English sounds"; stimuli were orally presented "only once." (Newman 1933: 62) In effect, each element (i.e. stimulus pair) was presented *twice* because in two separate sections of the experiment, subjects were asked to choose first the larger and secondly the darker[61] member of the pair. 226 word pairs[62] were presented; in order to minimize subject fatigue, the phonetic structure of the words were much varied: "Contrasted sounds were varied as to position in the word, accent and syllable structure." (Newman 1933: 63) which introduced several further variables into the already complex stimuli: Even though "variations were introduced consistently" into the artificial words (Newman 1933: 63), including them seems to have been an unnecessary complication. Reduction of the numbers of segments tested would have also minimized subject fatigue, permitted simpler stimuli and thus ensured more reliable responses.

Both for vowels and consonants, two different rankings were obtained for judgments of size and brightness. Newman's vowel scales are illustrated in Figure 2.2 below:

---

[60] For large/small, there were 141 subjects of whom 5 dropped out before the bright/dark test, leaving 136.

[61] Apparently, Newman did not consider it necessary to present both members of the antonymic term-pairs relating to brilliance and size. For discussion on the significance of this, see Chapter 7, Section 7.2.4.

[62] From the examples given, the words were minimal pairs but bisyllabic, which already introduced extra segments and therefore extra variables.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

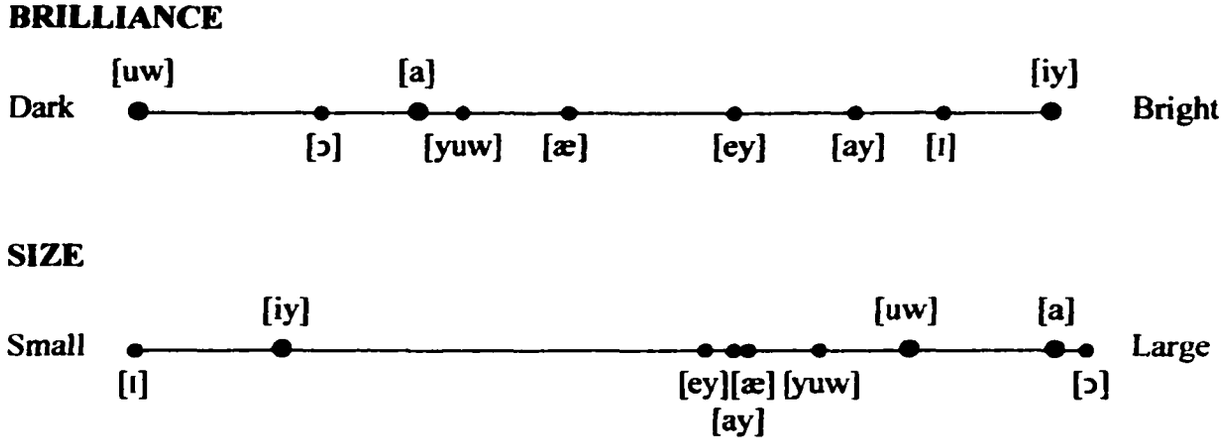


Figure 2.2, Newman's brilliance and size scales. Nine American English vowels scaled on brilliance and size dimensions; spacing is based on Newman's (1933: 63) findings. The four stimulus vowels of this experiment are enlarged and presented above the scale.

Newman noted that an influence in addition to the three factors previously mentioned was that of "vocalic length", noted in the 'misplacement' of [iy], surprisingly judged larger than [ɪ].[63] This is a similar finding to Sapir's with [e] and [a], and constitutes further evidence that in this kind of study the language-specific phonology of the subjects can be an essential influencing factor. It also indicates that perceived duration may compete with vowel height in its phonosymbolic effects, i.e., that more than one phonetic feature may trigger a symbolism and that the features may be in conflict.

It was observed that dark/bright judgments seemed to have been "patterned exclusively on an articulacy and frequency basis." (Newman 1933: 68) Newman decided that this must be because "factors of vocalic length and size-of-mouth opening evidently have nothing to do with the symbolic judgement of relative darkness, a situation which is entirely plausible."(Newman 1933:

---

[63] Newman (1933: 68): "The position of [iy] above [ɪ] on the present magnitude scale runs counter to both the articulatory and the frequency sequence. The conclusion can hardly be avoided that [iy] is symbolically larger than

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

68). However, examining the variations in judgments due to placement and accenting of the vowels led him to the following conclusions: "The symbolic value of accenting the vowel is heavily in favor of largeness and darkness; the final position of the vowel carries with it a large and dark symbolism." Newman (1933: 71) Since there is frequently a correlation between accenting and vowel length in English, perhaps the earlier conclusion was premature. What becomes clear from Newman's work is that the relations between the two symbolisms are not ones of inevitable entailment: *if dark then large, and vice versa* (as predicted by Tolman, von Hornbostel and others). This is further supported by the results of the judgments on consonants, which indicated that brightness and smallness (or darkness and largeness) did not always correlate: "The palatal consonants... are bright but large; the dentals, on the other hand, are dark but small." (Newman 1933: 69)

Following this second part of his study, Newman checked for possible lexical influences by collecting 500 size-related words from Roget's *Thesaurus*, concluding that "the phonetic content of English words takes practically no account of magnitude-symbolism." (Newman 1933: 74). His final conclusions were that "The basis of phonetic symbolism is fundamentally objective", that age "has little effect" on judgments, and that different symbolisms produce "diverse types of phonetic patterns" due to "a variety of objective factors", such as "the sequences of tongue position in articulation", "vocalic length," and so on. (Newman 1933: 75) Lexical influences, which he confusingly termed "linguistic associations" (Newman 1933: 75), were ruled out.

---

[i] because the length of the former vowel makes it 'sound' larger."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the same year, Bentley and Varon (1933) challenged Sapir's and Newman's findings as perhaps not due to phonetic symbolism at all, but to the more general phenomenon of sound symbolism. Five tests were devised, the first four of which gradually approached the design of Sapir's and Newman's; in the last two, an increasing number of subjects were used. However, the first three tests were of only "a small number of students trained in research" (Bentley and Varon 1933: 77), in fact, no more than three subjects. The first test was associational, using 40 series of 10 CVC word-sequences (made up of 2-3 naturally-occurring words and 7-8 artificial words), all orally presented. Under scrutiny was the contrast [a] : [i] as "critical" because "near the ends of Sapir's vocalic scale" (Bentley and Varon 1933: 77).[64] Subjects were asked to "respond promptly with the first sound[65] that came after hearing a given sound." No response 'sounds' were given "that suggested size or any determination of magnitude" (Bentley and Varon 1933: 78). This type of test resembles that of English (1916), cited above; given all the possible responses to a word, artificial or actual, it seems unlikely that a normal subject[66] would not respond with lexical associations well before any type of phonetic-symbolic ones.

In the second test, subjects were given 10 "category-words" (qualities, such as *large*, *angular*, *foolish*, *liquid* and *noisy*) and then an artificial word which contained either [a], [i] or English [ɛ]. Minimal pairs were used "sometimes".(Bentley and Varon 1933: 78) The 10 category-words were paired with 50 artificial words, so that each subject had 500 CVC syllables to judge as to whether

---

[64] It should be noted that the researchers were testing two non-English sounds.

[65] The term "sound" was used confusingly throughout this article to mean "word".

[66] That is any subject who was not a specialist in language, such as a language teacher, linguist, philologist, missionary, or literary person of some type.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the stimulus (artificial word) implied some degree of the category mentioned, or of its opposite. Conclusions were that "a little less than one-third of the whole number of reports" indicated a felt connection between any of the stimulus and category words. Instead, the authors suggested that the high degree of non-correlation "may be as significant as the *positive* reports." Moreover there was only a "slight relationship" reported between most stimuli and category words. (Bentley and Varon 1933: 79)

A third test was run with five categories only: "foolish, angular, liquid, large and strong", (Bentley and Varon 1933: 79) and with five vowels in the stimulus words: to [a], [i] and [ɛ] were added [ow] as in *go* and [uw] as in *tufa* (assumed here is that these were English long, tense dipthongized vowels, since they were exemplified in English words). This time, "in a still closer appraoch to the original method of Sapir," the catgory word was followed by two stimulus CVC syllables, and subjects were asked to choose "which (if either) of the two more closely approached the meaning of the category word just given." (Bentley and Varon 1933: 79) Each subject was assigned 100 coupled stimuli to judge, which appear, from the examples given, not to have been minimal pairs. Results showed little over half of the pairs reported as having one member related to the category word; the relationship was mainly "slight." (Bentley and Varon 1933: 80)

The authors concluded that Sapir's and Newman's previously obtained results were due to the fact that they had used forced-choice methods: their experiments had been designed "to force a meaning into a non-linguistic sound (i.e. artificial word) and then to suggest this meaning admitted of degree." (Bentley and Varon 1933: 80) Therefore, a fourth test was given, with 26 subjects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

They chose "three categories and their opposites (angular/round, large/small and hard/soft); partially forced meanings... we now used the 'arbitrary meaning' device of Sapir." (Bentley and Varon 1933: 80) Minimal pairs were used as stimuli. In an attempt to "remove visual influence,"[67] words were pronounced from behind a screen. Judgments once written down were removed immediately, so that they would not influence subsequent ones.

Results showed [a] was judged larger than [i] and [i] smaller than [a] in the relation 4 : 1 for the meaning "table". Similar results were found for roundness and softness, repsectively for the meaings "table' and "chair" with decreasing order of agreement: 3 : 1 and 2 : 1 respectively. However, the authors observed that "positive reports form only a moderate fraction of the whole" (Bentley and Varon 1933: 83). In their experiment, subjects were allowed to report on the *unrelatedness* of a category word to a stimulus pair as a way of controlling the forced-choice aspect of the design: "We have found that the qualifications suggested by 'arbitrary meanings' came out in proportion to the amount of forcing... The positive report did appear when *degree in some scale* was suggested and prescribed." (Bentley and Varon 1933: 83)

At this point in their article, Bentley and Varon suggested that any symbolism attributed to speech sounds was actually part of a more general sound symbolism,[68] the implication being that any

---

[67] Bentley and Varon criticized their predecessors for not controlling the factoring of responses "into auditory and kinesthetic components". In doing so, they make an important point about the complexity of speech sounds and the amount of potentially symbolic material they contain:"Five minutes' attention to the strains and postures incident to making brief vocal sounds will convince one that there is plenty of linguistic material which, given a linguistic turn, will easily absorb 'degrees' not only of magnitude but also of many other qualifications." (Bentley and Varon 1933: 81)

[68] "The simplest and best-ordered series of sounds (the tonal series) is a one-dimensional series, proceding from low, grave, voluminous, dull and weak members at one end to high, acute, small, bright and strong members at the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phonetic offshoot of sound symbolism was accidental and not properly linguistic. Their claim was that Sapir and Newman chose qualities (size and brightness) to investigate that are "exceedingly prominent in the simple tonal system"...but "pertain to the (speech) sounds *as sounds*. They attain a 'phonetic' or 'symbolic' signification only when we transfer them to other contexts" (Bentley and Varon 1933: 84—85).

Accordingly, a fifth test was done which mixed stimuli: three pairs of artificial words (CVC syllables, not minimal pairs), a pair of tuning fork tones and a pair of noises: a mallet thud vs. "the crack of a small pith hammer upon a wooden surface" (Bentley and Varon 1933: 85). The word-stimuli were compared for largeness, lightness and softness, the two tones for largeness and the two noises for smallness. Subjects were 327 university students. A greater unanimity of judgments were obtained for the tones and noises than for any of the CVC stimuli (tones 7:1, noises 20:1--but one word-pair did evoke a 6:1 response; the experimenters may have definitely inhibited responses on words by not using minimal pairs).

The authors concluded that many aspects of speech sounds could lend themselves to symbolic attribution, not simply the acoustic but the "visual perception... of the open mouth, stretched jaw, the tense lip," and the "tactual and kinesthetic component... the act of sound-production itself, which includes the feel of the tongue, lips, jaws, chest, and other parts of the body" (Bentley and

---

other... It is generally agreed among psychologists that some such set of aspects is actually attributive and therefore inherent in the tones, and that they gradually change from one end of the series to the other... There is in all these attributive qualifications a principle of surrogative function by way of which any one may stand for the other and may be interchangeably used." Bentley and Varon (1933: 83-4) So much for Newman's findings that brightness and smallness (and their opposites) did *not* automatically entail each other where certain speech sounds were concerned!

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Varon 1933: 86) However, they asserted that these made for symbolisms which were only incidentally linguistic.

Were Bentley and Varon correct in dismissing the idea of a special phonetic attributive faculty distinct from a more general sound symbolism? Their strongest argument against the experimental method used by Sapir and Newman seems to be that it put ideas (i.e. phonetic symbolism) into the subjects' heads. It relied for its effectiveness on first forcing a meaning onto an artificial word member of a minimal pair and then forcing attributions—a comparison of degrees of a quality (or its opposite) with the other pair member. The method was unsatisfactory, since it amounted to prompting the subjects into judgments that they would never have made spontaneously, in contrast with more 'natural' experimental tasks such as reporting first associations to stimulus words, as in Bentley and Varon's first test.

The arbitrary assignment of meaning to a phonetic string cannot seriously be objected to; possible cases of phonetic symbolism apart, this type of 'forcing' is the way word-assignment operates. But what of the forcing of attributions? In Sapir (1927) the point is made that though all language has a "marvellous completeness of formal development", nevertheless it "comes into use and conscousness merely as a general lever for all types of expression...of all the great systems of social patterning it is probably the one which is most definitely unconscious in its operation." (Sapir 1927: 421-3) The feeling of lexical fittingness comes within the realm of conscious awareness of the speaker because the object triggering it is an entire word: an entity which is easily detached from an utterance and thus easy to focus on. Recall that *Cratylus* (300 BC) begins with an exhaustive discussion of fittingness and only descends into phonetic mimesis after

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Socrates has led Hermogenes to a dead-end in trying to explain the fittingness of "a word which is incapable of further resolution...a primary element which need not be resolved any further" (*Cratylus* 422 b) i.e., one whose appropriatenss to its referent cannot be explained by literal meaning supported morphologically, or by metaphorical extension.

The problem, then, with testing for phonetic symbolism is that if it exists, since it deals in speech sounds and not in whole words, it is, like so much linguistic categorization, *an unconscious faculty*. The capacity to bring it into awareness varies greatly amongst individuals (the observations of writers like Guest (1838) Tolman (1887, 1904) Sherman (1893) and Smith (1912) are examples of such symbolizing intuitions made conscious, as are the reports by Alspach's (1917) single subject and Sapir's (1929) minimal-paradigm-forming subjects. However, it cannot be assumed that most subjects will have this kind of conscious capacity available. Therefore an experimental method is needed that will indeed force the subject (whatever his level of phono-symbolic awareness) to match speech sounds to qualities and judge to what extent they correspond.

Bentley and Varon (1933) are correct in requiring experimental design to include (as theirs did) an option for subjects to report *no relation* between a stimulus or stimulus pair and a quality; this is a guard against taking the forced-choice aspect too far and putting 'intuitions' into subjects' heads. Neither Sapir (1929) nor Newman (1933) described their procedures quite as fully as could be desired, but in the description of Sapir's second test there is a definite mention of providing subjects with a no-relation option (Sapir 1929: 229). In the first part of his investigation, Newman (1933) was using data from Sapir's second test; when he went on to construct his own

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

experiment, he reported that "the experimental and statistical procedure was *essentially the same* as that of Experiment 1,"i.e., Sapir's. (Newman 1933: 62, my emphasis.) It can therefore be argued that these early experimenters were from the outset aware both of the necessity for, and the perils of a forced-choice design: they guarded against the latter.

Bentley and Varon's other point (that the results of the earlier experimenters indicated only the presence of a general sound symbolism and not phonetic symbolism) brings us back to the remarks on the two terms made in Chapter 1. If a specifically phonetic version of sound symbolism exists, then it should both partake of that general symbolism and yet also in some area or areas clearly show the influence of the speaker's linguistic structure; one's native phonology or morphology should in certain instances affect symbolic judgments. It is surely significant that Sapir (1929) and Newman (1933) both separately mention results in which vowel length (or dipthongization, which is phonemic in English) seemed to be responsible for anomalous size-judgments. (Sapir 1929: 233-4, Newman 1933: 68). Bentley and Varon, however, made no reference to either of these reported anomalies.

In the same year as Bentley and Varon's, a very short, non-experimental paper was published (Tsuru and Fries, 1933) which nonetheless was to have a profound influence on subsequent experimental directions and methods. The authors provided 25 word pairs in English and the corresponding 25 pairs in Japanese, together with a key to correct matching of English-Japanese translations. Before using the key, the (presumably non-Japanese-speaking) reader was invited to match up the English pairs with the Japanese; it was predicted that "at least 75% of these pairs can be correctly guessed." (Tsuru and Fries 1933: 283) One obvious drawback to the test was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that only 80% of the pairs supplied were true antonyms, so that the capacity for intuiting meanings through contrasts could not be fully called into play. Another set of problems was posed by the complexity of using whole words as stimuli.

The authors claimed that correct word matching would be due to gestalt processes, and indeed the one reference cited was the gestalt psychologist Köhler (1929). They stated that the perception of a meaningful unknown word, the perception of a referent and the experience of a meaning are all conditioned by similar "organized processes" which have "specific causal properties of varying degrees of similarities" so that "the perception of that word and the thinking of the meaning will tend to give rise to organizations which have similar causal properties" (Tsuru and Fries 1933: 284). Since their paper was "a specific application of the theory that meanings, psychologically, are properties of natural organized wholes or processes" the authors were aiming their blow at Saussurean arbitrariness in the name of Gestalt psychology. In suggesting that meanings in one language were largely guessable by speakers of another, unrelated one, by implication they were also claiming the universality of phonetic symbolism.

1933 seems to have been an especially fruitful year for studies in phonetic symbolism. Newman had continued and developed Sapir's original enquiry, while Bentley and Varon's work raised important questions about experimental methodology and the linguistic or non-linguistic nature of sound symbolism. Tsuru and Fries chose 1933 to launch into the word-matching enterprise, and this was also the year in which Jespersen published his influential paper *On the symbolic value of the vowel I*, much cited in phonetic symbolism bibliographies ever since. It continued his moderate point of view, i.e. supporting the idea of phonetic symbolism but not dismissing Saussurean

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

arbitrariness. Jespersen proposed a drift or tendency in lexica (his examples drawn almost entirely from Indo European languages) to favor items (bound or free morphemes) which featured [iy] and [ɪ] linked to meanings of smallness.

The interplay between phono-symbolic influence and the lexicon can be seen as an absolute requirement, for example, that all or the majority of words and suffixes indicating smallness *should* contain [iy] or [ɪ]; this can be demonstrated not to hold true (Newman 1933). On the other hand, it can more plausibly be seen as a preservative and fortifying factor that "makes some words more fit to survive and gives them a considerable strength in their struggle for existence," part of a "slow, progressive tendency" in language "towards fuller, easier and more adequate expressions" using "an increasing number of sound symbolisms" (Jespersen 1933: 285-286).

In the same year, Grammont in his *Traité de phonétique* (1933))[69] also argued strongly for phonetic symbolism as one of the forces shaping the lexicon. His observations are highly detailed, reasonable and impressively exemplified (using French classical poetry), and this rarely-cited study merits an honored place in the literature. Basing his distinctions on articulatory facts, he divided the vowels in the traditional front-back way, the front ones being clear or bright (*claires*) and comprising [i] [y] [e] [ɛ] and [ø] (as in *feu*). Out of the clear group, those two articulated the farthest forward and pronounced "the most closed,", [i] and [y], he further distinguished as sharp or pointed (*aigues*). The group of back vowels were generally termed deep or hollow (*graves*), but divided into two subgroups, the brilliant (*éclatant*), comprised of [ɔ] and the [œ]of *peur* and

[69] It was republished in 1946; page references are from the later edition.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the dim or gloomy (*sombre*), grouping together the closed [o] of *clos* and the [u] of *mou* (Grammont 1946: 380—381).

Consonants, and more important still, the interplay between vowels and consonants, were treated in a similarly precise manner, with the vowels seen as primary and the consonants as modifying their effects. Thus, Grammont is one phonetician who perfectly understood the implications of the fact that words are composed of several segments: any symbolism conveyed by a word "is complex, and the various elements which contribute to its production are combined with each other, react with each other, strengthen or weaken each other in such a way that sometimes in order to disengage one of them, one has to lean on the facts of general phonology" (Grammont 1946: 391). Such a view is very far from that of the word-matching experimenters, who seldom if ever thought to subject their word-pair stimuli to any form of phonetic analysis.

Having dealt with literal, onomatopoeic effects as primary in phonetic symbolism, the *Traité* also included an important section on the metaphoric extension of phonosymbolic effects. Grammont made the distinction between *imitation* (onomatopoeia) and *translation* (expressive, metaphoric use of speech sounds, true phonetic symbolism): "Sharp vowels, when expressing sharp sounds, are not translating, they are imitating; but through translation they can give the impression of the sharpness of an object, as in the word '*aigu*' [sharp] itself... or moral or intellectual sharpness... Since the sharp vowels penetrate our ear in the same way as a sharpened blade and sometimes give us a neighboring expression of pain, they highlight a certain number of words which express sadness or horror and which are like a cry" (Grammont 1946: 404).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Grammont also noted a phenomenon occurring in poetry that Macdermott (1946) was to observe a decade or so later: the patterning of vowel-groups in verse. "It is the ear and mind which quite unconsciously groups the vowels and compares the groups... The clearest modulation (i.e. from one group to another) is produced by passing from a hollow vowel to a sharp one or from a sharp to a hollow" (Grammont 1946: 421). However, his most important observations are on the symbolic interplay between vowels and consonants in words, and on the notion of a central onomatopoeic sense of imitation as the core starting-place for all penumbral phonosymbolic translations.

Thus, as a watershed year, 1933 saw five important publications representing different streams of thought in the general flow of work in the area of phonetic symbolism. In this chapter, the heterogeneous progression of writing and experiment on phonetic symbolism (now philosophical, now literary, or philological, linguistic, psychological) has been treated as a unitary, chronologically charted, flow until this point. But from 1933 to the present, it will be clearer to continue the review by tracing three lines of enquiry. Chapter 3 focuses on lexical and poetical studies; language itself is being studied here, and not the responses of a group of subjects. Research in this area attempts to locate "objective phonetic symbolism" (Taylor and Taylor 1965, Mandelker 1983: 12) within the lexica of a language or group of languages. Since it has frequently been claimed that objective phonetic symbolism is noticeably employed in the sound-patterning of the poetry of a language, a few significant studies of poetry have also been included, though there is no claim to a comprehensive survey. As with synesthesia studies, this is a vast field; only what seems of especial relevance has been inlcuded.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 4 explores studies of fittingness and synesthesia, covering both a small amount of theoretical work, and some fittingness or synesthesia experiments, often using figures and drawings which are frequently but not always matched to words. The experimental focus is on synesthetic correspondances as a function of gestalt organization. Because the investigation of synesthesia is an enormous field in itself, only those studies of direct relevance to phonetic symbolism are mentioned.

Chapter 5 covers a number of different areas of linguistic experiment, one area being word matching studies. Begun in 1933, with Tsuru and Fries' short report, this type of experiment became enormously popular during the late 50's and the 60's when Chomskyian obsession with universals entered the intellectual scene and any chance of proving their existence was eagerly siezed. It is included here because of its historical importance, and as an example of the kind of experimental blind alley that misapplication of the concepts of phonetic symbolism can lead us into. A second area is minimal pair experiments. The focus here is on stimuli varying one speech sound or feature at a time, with the aim of pinning down segmental and feature symbolism. I will argue that, given everything which we *don't* know about phonetic symbolism, a minimal, step-by-step approach like this is a reasonable way to investigate connection between speech sounds and qualities. Under this heading come studies like Sapir's and Newman's that use artificial words, and later studies like Tarte's (1982) which use single speech sounds. My own experiment is of this type. Minimal pair studies using stretches of discourse, such as Hevner (1937) and Stevens, De Vito and Isaacson (1969), are also included.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 3

PHONETIC SYMBOLISM: LEXICAL AND POETICAL STUDIES FROM 1934 TO THE PRESENT

3.1 General introduction

On looking over the investigation of phonetic symbolism naturally occurring in language[1] up till the present, certain themes emerge. The most obvious is the quest for systematicity—to determine the degree of phonosymbolic orderlinness in the lexicon, and to understand (in a general sense, extending beyond phonetics/phonology to morphology and syntax) how far all language is pervaded by iconicity. A related issue is whether phonetic symbolism (and linguistic iconicity in general) is an either/or phenomenon, like voicing or nasalization, or should be located on a more-or-less continuum, like syllabicity or obstruence.

One type of systematic investigation examines the traditionally-acknowledged phonosymbolic qualities of size and brightness, and tries to determine how far words naming those qualities will contain the same sounds, notably the same vowels. Another approach looks at conceptual relationships, such as proximity and distance, and examines phoneme distribution in words that denote these types of relations. And, as lexical cross-

[1] This is sometimes referred to as "objective phonetic symbolism," cf. "subjective phonetic symbolism," i.e., specific phonosymbolism attributions by subjects to linguistic material (see Taylor and Taylor 1965)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

influences become a concern, there is debate over how far all word-surveys should be diachronic or synchronic.

Begun with Jespersen's (1922) groupings, there is also an ongoing discussion of what constitutes a phonosymbolic class, beyond that of the most obvious type of onomatopoeia. The concepts of the *phonestheme* (Firth, 1930; Householder, 1946), and of lexical *constellations* (first described by Bloomfield, 1985; see also Bolinger, 1956; de Saussure 1956) bound together by common phonesthemes, gradually move from being marginal to being respectable and serviceable in explaining linguistic structures. Amongst the ideas proposed, perhaps the most important in English is lexically-generated *secondary phonetic symbolism* (Bolinger, 1965). This is described and acknowledged as a shaping force in the lexicon and as an extra cross-referential resource in lexical retrieval.

Related to the above issues is the debate over how the word should be divided. What, if any, is the dominant element of phonetic symbolism in a word: is it onset, nucleus, coda—or rime, which encompasses nucleus and coda in one unit? Wherever I have found reflections on this topic—not normally considered crucial—I have gathered and included them. They, together with more synesthetically-focused remarks (such as those of Werner and Kaplan, 1963) constitute conceptual groundwork for the later section on word-matching studies.

Another line of exploration takes poetry, with its focus on expressiveness and its marked degree of phonetic structuring, as having an especially high concentration of phonetic

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

symbolism—like normal language, only more so. Thus, the poem is seen as providing excellent laboratory specimens for the linguist, a view which starts with Tolman (1887, 1904), and perhaps farther back, and which is also brilliantly exemplified in Grammont (1933). A linguistic approach to poetic analysis consequently becomes a useful way both of understanding hitherto mysterious expressive-esthetic effects and of discovering general phonosymbolic patterns.[2]

Finally, and perhaps most important for my own work, is the question of universality—how far different languages can be said to have common, predictable patterns of phonetic symbolism. In the same way that I have been searching for remarks about phonosymbolic cross-influences in words, I have taken note of all reflections touching on the issue of the possible multiple symbolisms of speech sounds (both within one language and across languages) and tried to highlight them in this section.

3.2 Studies.

The first writer[3] to be treated here is Orr (1944), together with the response to his article by Thorndike (1945). Orr further explored the traditional observation that high front vowels are associated with smallness and low back ones with largeness, which contrast he termed "vowel antiphony" (Orr 1944: 3). He cited a large number of minimal or near-

[2] For a somewhat different view, see Brown (1958: 138—139) who concedes that phonetic symbolism may be found in poetry, but calls it "a non-laboratory situation," and cautions against expecting simple onomatopoeia: "I do not doubt it [phonetic symbolism] ordinarily assumes a subtler form."

[3] Although Hevner (1936) chose to examine poetry in her study and predates Orr, given that the stimuli of her main experiment were in the form of poetical minimal pairs, she has been somewhat reluctantly classified in Chapter 5.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

minimal English word pairs, which all employed this symbolism,[4] and pointed out that he could not cite any pairs of counter-examples, let alone give a list. He quoted further instances to show that the expressive sequence front-vowel-to-back-vowel in a string of syllables or words seemed to be favored in French and English[5] for kinesthetic or auditory reasons—either "because of the tension and release experience in a sequence like /i/-/a/ or because there is a sense of climax in finishing a word or a sentence with a sound of substantial volume" (Orr 1944: 6).

Orr concluded with a warning of the "ever-present danger of self-persuasion and subjectivity in the matter of symbolic representation," yet he still maintained that "there is a feeling of satisfaction in the utterance of a 'front-back' vowel sequence... *the difference in tension and sonority has also a symbolic, one might almost say pictorial value, due to some process of psychological transference*" (Orr 1944: 6, emphasis added).[6]

Inspired by Orr's observations, Thorndike (1945) responded by making a list of English monosyllables containing /ɪ/ or /iy/ "that occur once or more per million running words" (Thorndike 1945: 10) and inspecting them for meanings of smallness and largeness. As a control, he similarly inspected a list of monosllyables with the same initial and final consonants but containing all other English vowel sounds[7] "The facts," he remarked, "give

[4] Such as "slit, slot; slipp(er)y, sloppy; chip, chop;"and so on (Orr 1944: 5)

[5] Among examples given were French "ad-mi-rable" and English "safe and sound."

[6] Even in this purely lexical study, the author noted a synesthetic effect; this could be an influence on word formation and thus on some small part of the structure of the lexicon.

[7] With the exception of /ɔy/ the rhotic diphthongs, and of course schwa.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

strong support[8] to Professor Orr's hypothesis" (Thorndike 1945: 10). However, since out of the high front group, the actual percentage of /ɪ/ or /iy/ monosyllables meaning "small" was found to be 0.07%, and out of the low back group those meaning "large" to be 0.05%, "the difference is small enough that, as far as I can ascertain, no psychologist or linguist until Professor Orr had tried to prove its existence." (Thorndike 1945: 11).

Thorndike then cautioned against the unthinking universalism of assuming that the front-back contrast would tend to have the same meaning influence on the lexica of other languages.[9] Small data samples from Greek, Hungarian and Finnish were offered in support of this statement; the author concluded that in these languages "the front and back vowels... do differ in their suggestiveness of smallness and largeness, but the differences are much less than in modern English... I am under the impression that the diffferences [in Old English] are much less than in modern English" (Thorndike 1945: 13).

Thorndike's lexical study contrasted in a couple of ways with that of Newman (1933: 73), who included polysyllables in his survey and drew his word sample from those sections relating to size in Roget's *Thesaurus*. Newman thus included both exotic words, such as "pigwidgeon," and words only tangentially related to size, such as "nutshell," "scrubby", and "tincture." Both surveys seem partial at best, Newman's because his source material

---

[8] "The smallness to largeness ratio is 12 times as large for /ɪ / and /iy/ as it is for /ɔ/ and /ow/ and there is no likelihood that this difference is due to chance" (Thorndike 1945: 10).

[9] "We should canvas the significance of the front versus the back vowels in several widely separated linguistic stocks before concluding that it is largely a matter of common human physiology and psychology. Such measurements as I have had time to make seem, indeed, to show that *English is a rather special case in the relation of front and back vowels to smallness and largeness respectively*" (Thorndike

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

was so heterogeneous and was not controlled for frequency of occurrence, and Thorndike's because in taking the simpler course of treating only monosyllables, he excluded a large amount of data and possibly skewed his study.[10]

According to Householder (1946) the small/large contrast of the high front to the low back vowels is the opposition of two *phonesthemes*[11], a phonestheme being defined as "a phoneme or cluster of phonemes shared by a group of words that also have in common some element of meaning or function, tho' the words may be etymologically unrelated" (Householder 1946: 83). Like Thorndike, Householder confined himself to monosyllables, but his study was focused on only one vowel: /ʌ/. He collected 266 standard English and 400 dialect words containing this vowel, and sorted them into groups with 9 general meanings.[12]

In a brief report on a paper describing his investigation, it was remarked that "besides the frequently recurring pejorative force, there seem to be two or three specific senses belonging to this phonestheme, but *they cannot be precisely defined until due allowance has been made for the various consonant phonesthemes present in these words and some*

1945: 12. emphasis added).

[10] For example, in counting Newman's "smallness" sample (Newman 1933: 73) one finds 183 words, of which only 72—just over a third—are monosyllables. Newman's monosyllables also include exotic and semantically tangential examples such as "gnat" and "whit," so it cannot be claimed that limiting a survey to monosyllables would necessarily produce more 'basic' vocabulary.

[11] Wescott gives the exact origin of this term, as "introduced by J.R. Firth, as in *Speech* (1930; rpt London: Oxford Univ. Press, 1966; p 187)" (Wescott 1971a: 129, fn)

[12] Meaning categories included "adverse judgement... something cut off or broken off... (words) referring to making a dull, somewhat loud but indistinct noise... (words) referring to being a coward or a failure or

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

study of the contrasting vowel phonesthemes" (Householder 1946: 84, emphasis added).[13]

Following this report, criticism of Householder's method was also recorded by Fowkes who stated that "sound methodology is violated by assigning subjective, synchronic significance to a phenomenon which is actually diachronically determined" (Householder 1946: 84). Fowkes' was precisely the point that Jespersen argued against,[14] voicing the logical, Saussurean outlook that refuses to credit any phonosymbolic influence in the forming of words or of the shaping of the lexicon.

Macdermott (1946) noted a phenomenon occurring in poetry that Grammont (1933) had earlier observed: the patterning of vowel-groups in verse. "It is the ear and mind which quite unconsciously groups the vowels and compares the groups... The clearest modulation (i.e. from one group to another) is produced by passing from a hollow vowel to a sharp one, or from a sharp to a hollow" (Grammont 1933, 1946: 421). Thus, according to Grammont, the most salient vocalic contrast, and therefore the most useful poetically, is that of high front to low back.[15]

---

to deceiving... strong verbs, meaning past tense or past participle or both" (Householder 1946: 83).

[13] This is the first allusion, apart from Grammont (1933) that I have found to the possible complexity of phonetic symbolism in words, and the consequent the need to take all segmental elements into account when trying to pin down which might be responsible for especial fittingness of meaning. It also underlines once again the need to think in terms of *oppositions* when looking at symbolisms.

[14] See Chapter 2.

[15] But see Fischer-Jorgensen's comments on Westermann and rounding as *the* crucial phonosymbolic feature (Fischer-Jorgensen 1978: 85), also Wescott (1971b: 420) on the three chief axes of vowel polarity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the *Traité*, this phenomenon was described as occurring in a single line; Macdermott's discovery, examining classical English poetry, was that sections comprising several lines would be dominated by one or the other of the two sets of back or front vowels,[16] forming what she described as "units" (Macdermott 1946: 14). Her study examined 250 passages of poetry phonemically transcribed containing single, double or triple units, contrasting good and bad poems and focusing on what she termed "tone-color" effects with particular reference to back-front vowel patterning. Good poems were found to have vowel patterning that corresponded to their content; bad poems lacked such patterning or employed it inappropriately for their subject matter.

As Macdermott continued her study, she came to the conclusion that besides the grouping of vowels as back or front, there were two further necessary categorizations to be made. One was the "deep group," composed of /uw/, /U/, /iy/ and /i/; the other was the "common group," a mixture of low front and mid and low central vowels: /ae/, /a/, /ʌ/ and /ɘ/ (Macdermott 1946: 77-82). Her evidence for the validity of these groupings was based on the acoustic studies of vowels to date, especially early work on formants. Her hypothesis was that the ear could pick out and associate a formant or formants belonging to one vowel with that or those of another, as if the undernotes of two chords could be similarly associated by a (musically uneducated) ear, and thus classed as belonging to the same group.

---

[16] This may be highlighting an actual difference between the use of vowel patterning in French and English poetry.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Dr. Larry Raphael has advised me (in a personal communication) that the classification of vowels in the deep group has some acoustic plausibility, since all the vowels share a low F_1. However, the common group (which supposedly could ally itself with either the back or front character of a poetical unit) he considers too hybrid to be a convincing acoustic category.

Macdermott's study, though full of interesting material, is often perplexing to follow. Despite coming sixty years later, it strongly recalls the work of Tolman (1887, 1904). She arrived at very similar conclusions to his about the poetic application of back and front vowels to oppositions in qualities of motion, direction, perfumes, types of light, states of mind, times of day and so on.[17] Like Tolman, she also felt that the intrinsic pitch of vowels was a significant part of their symbolic identity; however, this led to her calling the back vowel group "low" and the front group "high," one of the minor confusions to be found in her book.

Several years later, a much smaller but influential phonemic analysis of some poems was made by Lynch (1953), apparently under the impression that such a thing had "never before [been] attempted for English poetry." (Lynch 1953: 211). Lynch's assumption was that examination of the phonemic structure of a poem would reveal "how the poem's phonemic totality supports... its prose and poetic statement" (Lynch 1953: 211). His central specimen (Keats' *On First Looking Into Chapman's Homer*) was transcribed

---

[17] For a detailed list of such oppositions, see Macdermott (1946: 89-90)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

synchronically, i.e., without reference to its actual historical pronunciation.[18] He found that /ay/ and /ə/ were the dominant vowels and that /n/, /d/, /t/, /l/, /t/ and /s/ the dominant consonants. After making a literary analysis of the poem's meaning, he concluded that "silent" was the word which both "sums up the theme of the sonnet" and "also sums up its dominant sound structure" (Lynch 1953: 219).[19]

A similar conclusion was reached for Wordsworth's *I Wandered Lonely As A Cloud*; the change in imagery and tonality in the last stanza of Arnold's *Dover Beach* was also found to be marked by a change in dominant phonemes, but unfortunately there was no attempt made to analyze dominant phonemes in terms of articulatory or acoustic features. Lynch compared frequency of phonemic distribution in the Keats sonnet with two distributional studies of phonemes in written and spoken prose; notable differences emerged in the comparison of vowel distributions, but once again there was no attempt to explain them. The author claimed for his method no more than that it might aid in "getting closer to the total meaning of other poems" (Lynch 1953: 223). However, his discovery that a thematically central word or words in a lyric poem might act as hologrammatic keys to the poem's sound-structure is suggestive and radically different from the poetical discoveries of Hevner (1936) and Macdermott (1946).

Lynch's discoveries were dismissed by Chatman (1957) in a somewhat contradictory article. Following Bentley and Varon (1933), he criticized Sapir's (1929) and Newman's

---

[18] The author gave no reasons for this.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(1933) research for its use of "forced response techniques" (Chatman 1957: 249) and repeated Newman's confusion of hypothesizing (but not finding) an /a/-/iy/ magnitude symbolism embodied in size words in the English lexicon.[20] Chatman was very cautious in recommending the possible applications of phonetic study to poetry; he reasonably claimed that frequency analyses of segments in poems were useless without a clear control (i.e., a reliable measure of segmental frequencies in normal speech). Thus, he dismissed Lynch's discovery of /s/ /l/ /n/ and /t/ as being saliently dominant phonemes in Keats' *Chapman* sonnet on the grounds that these phonemes are ones that will be the most frequent "in *any* extended text" (Chatman 1957: 250).[21]

However, he conceded that where a poem displays a strong emotional or esthetic effect, frequency counts of phonemes or consonant clusters which show clear divergence from normal distribution may be useful. Examining marked or salient phoneme distribution may be able to tell us something of how the poet produces a "preformed impression in the sensitive reader's mind" (Chatman 1957: 250). This seems a somewhat backhanded admission that phonetic symbolism studies might have something to teach the poetry critic after all.

---

[19] For an almost word-for-word description of the same phenomenon, see Grammont (1946: 404).

[20] For a convincing justification of forced-choice experimental design and explanation of why one will not encounter tidy, total areas of phonetic symbolism in the lexicon, see Werner and Kaplan (1963: 218—219)

[21] In support of this observation , see also Fudge (1970: 171, 177) on the unmarkedness of English alveolar consonants.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the following year, quoting Jespersen's (1933) work as an inspiration, Maxime Chastaing published a detailed, two-part article on the symbolism of /i/, drawing mainly on the French lexicon (Chastaing 1958). Grammont was also clearly an influence on his thinking.[22] The first section of his article dealt with how /i/'s become "acoustic onomatopoeias" and "kinetic," while the second section looked at /i/'s which are "symbolic of sharpness and smallness." (Chastaing 1958: 405). Clearly, he was following Grammont's (1933) distinction between literal onomatopoeia (imitations)—the core—which he treated first, and symbolic applications (translations)—the penumbra—which were discussed later. However, Chastaing's terminology was more confusing than Grammont's, since his primary term for fitting words was "*onomatopées*" (onomatopes). (Chastaing 1958: 403), while other synonymous terms were "*symboles*" and "*signifiants motivées*." Thus the literal meaning of onomatopoeia, so essential to Grammont's argument, was lost.

The most important feature of his article, which is full of suggestive observations, was the Jespersonian integration of de Saussure's ideas of oppositions and linguistic structure with an acceptance of the workings of phonetic symbolism. In true Saussurean form, he observed that even the most onomatopoeic utterance is still a linguistic construct: "Verbal imitation means neither imitation nor 'approximate' copy, but construction and interpretation" (Chastaing 1958: 404). Chastaing also maintained that symbolism will be language-specific, influenced by every language's individual phonemic configuration: "The

---

[22] No less than four works by Grammont are cited in his bibliography.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

form of onomatopes depends... on the phonemes available to each language..." The vowel /i/ is "'the highest note'...given the orchestration of my language... the formerly absolute dignity of /i/ now appears relative. Dignity of the distinction: neither /u/ nor /a/" (Chastaing 1958: 404).

His article also continued Jespersen's (1933) and Orr's (1944) practice of listing phonosymbolic oppositions (such as /i/-small and /a/-large) that are lexically embodied in language.[23] A word "...is motivated by words whose sounds differ from it and by what those words mean" (Chastaing 1958: 413). He noted how the "physiological triangle" formed by the point vowels in French lends itself to minimal pairs: "*siffler-souffler, grigner-grogner*" (Chastaing 1958: 417). Quoting Edith Sitwell's line "The trees are hissing like green geese,"[24] he commented "A motivated sentence seems to us like a proposition that has been phonetically confirmed" (Chastaing 1958: 409). This observation looks forward to Werner and Kaplan's (1963) concept of *physiognomization*, in which the basis of meaning is found in the link between the physiognomized perceptions of both a referent and of the utterance that denotes it. Poetry (especially work like Sitwell's) loosens the ordinary referential (Sapir: 1927) use of language in order to focus on physiognomizations: in this case, the sight and sound of leafy trees agitated by wind.

---

[23] Chastaing mostly used French examples, but quoted from many other Indo-European languages, especially English.

[24] Sitwell was an English poet, known for her technique of foregrounding phonetics at the expense of discursive meaning in order to create emotional impressions. Her poetry relies a lot on synesthetic effects. She wrote in a personal communication to Maurice Bowra, the classical scholar: "Sometimes when I begin a poem, it is almost like automatic writing. Then I use my mind on it afterwards."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the second section of his article, that is, the one concerned with phonosymbolic translations to use Grammont's term, he dealt with the synesthetic linkage of senses which underlies all applications of phonetic symbolism: "/i/'s correspond on the one hand to sounds that excite us; on the other hand that we connect with tactile stimuli. Why this connection? Because we lose, when we feel excited, that Cartesian coolheadedness which allows us to distinguish between the gifts of touch and the fruits of hearing. Because we confuse even more easily these with those in that they are frequently associated" (Chastaing 1958: 461—462). Moreover, he included typographic aspect of /i/. Noting its shape, he remarked that the letter representing it is sharp, "resembles a finger pointing; [as a] phoneme, it designates pointed forms; typographically it draws them." (Chastaing 1958: 462).

The second part of the article also traced such /i/-associated symbolisms as those of excitation, acute pain, laughter, tickling and response to tickling. Closeness was yet another phonosymbolic translation: "/i/... because of its advanced position, seems to our interlocutors closer to them than to other vowels" (Chastaing 1958: 478). In a final exhortation to continue investigating /i/, Chastaing mentioned "studying the variations in the symbolism of /i/ depending on the variations of the surrounding consonants" (Chastaing 1958: 480), which has a very Grammontian ring. In any case, it showed that he was fully aware of the possibility of segmental cross-influences in words.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hans Marchand also published a very extensive article on phonetic symbolism (1959) from the point of view of the scholar of word formation.[25] Like Grammont, he distinguished two types of symbolism (or more properly two symbolic strategies): to imitate "things which we percieve through our senses (*direct imitation*)" and "to express feelings (*expressive symbolism*)" (Marchand 1959: 146). At first sight these seem close to Grammont's classifications, but on examination it turns out that the first is wider and the second narrower. "Direct imitation" appears similar in conception to Grammont's "imitation," but whereas Grammont referred *only* to onomatopoeia, Marchand extended it: "As noises and sounds are often accompanied by movements (as in *whish*, *swish*, *dash*, *tap* etc.) these also come to be denoted by symbols.[26] By extension, even the originator of a sound may be characterized by the use of a symbol (e.g. *pom-pom*, 'kind of machine gun')" (Marchand 1959: 146).

On the other hand, the concept of expressive symbolism was treated in a restricted way: as belonging only to words that directly express emotions, such as diminutives and contemptuous interjections. Marchand did not consider the emotional-attitudinal haloing of words denoting objects, actions or qualities (as, for example in *cool*, *zit*, *flummox*, *nasty*, *humungous* and so forth.) as phonosymbolically analyzable. With regards to the possible wider application of expressive symbolism, as in the above examples, he declared

---

[25] This paper was followed ten years later by a section in his book *English Word Formation* (1969) which repeated the article more or less word for word. It will therefore not be treated here. I am indebted for the original (1969) reference to Professor Charles Cairns—Marchand, like so many writers on phonetic symbolism, has not been accorded his due place in the literature and is all too rarely cited.

[26] This recalls, once again, Darwin's (1955) principle of serviceable associated habits.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that the problem of the origin of phonetic symbolism is "one of general linguistics" characterized by "the impossibility so far of finding out what the symbolism is based on" (Marchand 1959: 147).

A discussion of the use of phonetic symbolism in English words then followed, together with an impressive survey of examples. He remarked significantly on the phonosymbolic interplay between nucleii and codae,[27] "In direct sound imitation, the vowel denotes the pitch, volume and timbre of the imitated sound. *But whether the sound is sharp, cutting, vibrating etc. is indicated by the following consonants*" (Marchand 1959: 150, emphasis added).[28] This has the ring of direct observation, of a man much used to noticing different types of lexical nature, behavior and effects. Speaking as a scholar of word formation, going beyond his earlier rather limited classifications, he asserted the need to know "which particular phonemes are used in a language for the coining of words and what the symbolic value of the respective phoneme is. This will lead us to the question of whether a certain idea may not necessarily call for a certain sound" (Marchand 1959: 152).

Continuing this line of thought, he argued against too-easy overgeneralization. Some sounds may have universal symbolic application, but "there are other elements which are different in each language. As every language has its own phonological system, onomatopoeic coining is largely dependent on the phoneme combinations of the language"

---

[27] The section on individual vowels (pp 150—151) contains very little—apart from details—which has not already been treated by other authors; I have therefore not included a description here.

[28] For further focus on coda symbolism, see Wescott (1971a).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(Marchand 1959: 153). This remark also has an obviously functional ring; parallells can be seen in the language-specific techniques of poetry, which, like word coining, is a creative process. Thus, just as the Alexandrian poet Constantine Cavafy (1863—1933) made use of the presence of Greek diglossia as an expressive resource in his work, or the English poet Gerard Manley Hopkins (1844—1889) used the elasticity and contractability of English vowels or the double Latin and Germanic vocabulary of English in his, so in word formation, the general creative faculty of a linguistic community seizes on whatever its language possesses (in this case, its phonemic inventory, phonotactic rules) as a resource to express itself. Whether individual or collective, the creative principle is the same.

Marchand's phonosymbolic analysis of words was bi-morphemic, breaking them down into onset and rime. "Only two morphemic elements... play a relevant part: the initial symbol, i.e. the consonant(s) preceding the vowel, and the final symbol, i.e. the vowel and the final consonant(s)" (Marchand 1959: 153-154). He appeared to be implying that onsets are symbolically dominant: "To deny the word-forming character of initial symbols is hardly possible... If we consider how many words suggestive of the same idea are characterized by the same symbol, it is impossible to deny that the coinage was prompted by the symbol" (Marchand 1959: 157). However, rime too could be influential: "Its part (in the derivation of symbolic words) is obviously more important than has hitherto been recognized" (Marchand 1959: 265).

Reasserting the principle of arbitrariness, he also noted that not only primary imitative sound-meaning associations but even unmotivated ones can generate the forming of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

semantic paradigms linked by, e.g a common onset: "Most simple and all compound symbols are nothing a priori, but have developed morphemic character from the more or less accidental grouping of semantically related words. In *fl-* there is nothing to suggest flying or flowing movement, but in the co-existence of *flow*, *fleet*, *flutter*, *fly*, *flee*, *float* (all OE) lies the germ of all the new words expressive of movement which were coined in ME" (Marchand 1959: 154).

This is an important point, because it is being argued that paradigmatic formations in turn generate a type of secondary phonetic symbolism[29] Hamano (1986) indicates that it is this secondary lexical-paradigmatic principle which is also at work in the coining of new *gitaigo* (expressives or attitudinals) in the Japanese lexicon. Since these words form a separate syntactic class (as they do in Korean), rather than the English case of many series of lexical clusterings or constellations (as in Marchand's *fl-* example), perhaps the tendency to operate with predictable phonetic rules in word coining could be that much stronger in Japanese or Korean than in English. On the other hand, informal but very prevalent clustering of the English type may have exactly the same reinforcing effect.[30]

In the same year as Marchand's, a short study was published by Murdock (1959) validating the claim that *ma/na* was the most common word for "mother" in the world's languages and *pa/ta* the most common for "father. He surveyed kin terms in 474 societies,

---

[29] Bolinger (1965: 183—267) treats this topic in English extensively and convincingly.

[30] For the suggestion that sound-meaning constellations may indeed play a productive part in English-based phonetic symbolism, see Brown (1994).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

classified into 49 distinct linguistic families to avoid overweighting due to too many examples of three dominant groups. To make the survey, it was necessary to "reduce the immense phonetic variety of the actual kin terms to a number of basic sound patterns" (Murdock 1959: 2). Thus, vowels were reduced to the triangle *i-u-a*, consonants to 13 types and polysyllables to the root syllable.[31] In Murdock's survey of simplified kin terms, basic oppositions (e.g. nasal versus oral consonants, high versus low vowels) appear to be the motivators for contrasting mother and father terms: "(father) will tend to be called by a combination of sounds that is nearly as easy to utter and at the same time presents a clear contrast to the nursery word for mother" (Murdock 1959: 1). However, this conclusion may be due to schematization of the data—which is the problem attendant on this kind of reduction.

Perhaps the most interesting point in this paper is that Murdock was explicitly surveying *infantile* forms of the kin terms—what Werner & Kaplan (1963) call "physiognomized language": "As standard parental terms become phonologically and morphologically modified in consequence of the normal processes of linguistic change, forms develop which are very difficult for young children to pronounce.[32] Under such circumstances, simpler nursery words tend to appear—carved, so to speak, out of infant babblings under parental encouragement." (Murdock 1959: 1). However, these forms are clearly not

[31] This recalls Iritani's (1962) unpublished experiment quoted by Werner and Kaplan (1963), in which to analyze associations to a variety of artificial words, Iritani reduced his stimuli to to 6 polarized dimensions. He then found that apparently random free associations were varieties of response to the same semantic feature.

[32] Werner and Kaplan (1963) maintain that children are not *cognitively* ready (as well as not articulatorily ready) for the use of adult forms.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

onomatopes (imitations), the most primitive type of symbolic utterance, according to Grammont (1933) and Werner and Kaplan (1963) but penumbral, transferred translations. Even so, if there were any place to find phonosymbolic universals, i.e. symbolisms relatively uninfluenced by morphological and lexical material, this kind of infantile, essential vocabulary would be one good place to start.

In a long article arguing the need for linguistics in the study of poetry, Jakobson (1960) made two important points. The first eloquently defended the existence of phonetic symbolism as "an undeniably objective relation founded on a phenomenal connection between different sensory modes, in particular between the visual and auditory experience." (Jakobson 1960: 372). Jakobson put his finger on the methodological drawbacks of much phonosymbolic experimentation, for example, word-matching studies: "Particularly from the linguistic point of view, the picture has often been distorted by lack of attention to the phonological aspect of speech sounds or by inevitably vain operations with complex phonemic units instead of with their ultimate components. But when on testing, for example, such phonemic oppositions as grave versus acute, we ask whether /i/ or /u/ is darker, some of the subjects may respond that this question makes no sense to them, but hardly one will state that /i/ is the darker of the two" (Jakobson 1960: 372—373).

His second remark echoes that of Marchand quoted above on the language-specific shaping of expressive utterance due to phonology: "An analysis of poetic sound texture must consistently take into account the phonological structure of the given language, and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

beside the overall code, the hierarchy of phonological distinctions in the given poetic convention" (Jakobson 1960: 374). Poets work with what is to hand linguistically; of course, this will vary from language to language. The same point needs to be made about phonetic symbolism, both in onomatopoeic and expressive word coining and in any other manifestations (such as attributions to experimental stimuli by different language groups). Surprisingly, Jakobson did not mention lexical and morphological influences (i.e. types of secondary phonosymbolic triggers, just as much part of the language as its phonology) as also responsible for phonetic symbolism.

What is important here is to balance Jakobson's assertions; there may well be some kind of absolute for all speakers in which the /i/-/u/ contrast will always be perceived as respectively brighter/darker, and yet in the specific phonological structure of a given language, /i/ may *not* be perceived as the as the brightest sound.

Another article published in the same book as Jakobson's (Hymes 1960) took Lynch's method of phonemic analysis of poetry as a way of firstly examining "orchestration" (the total effect of a poem's sound quality) and secondly of relating these findings to the meaning of the poem. Following Lynch's (1953) discovery of the word *silent* as being the thematic and phonetic emblem of a Keats sonnet, Hymes specifically searched 20 sonnets for words which would contain dominant or very frequent speech sounds, would express the theme of the sonnet and would be placed late for a culminating effect ("summative words").

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hymes criticized Lynch for not contextualizing his results (not looking at sound frequencies in other poems or works by the same author, or general sound distributions in language) so he examined the frequency of sounds and their weighting[33] (absolute values and rank order, i.e. prose stress and metrical stress). Out of twenty sonnets, he found that six had summative words, five, key words (lacking late placement but with the other two characteristics) while seven displayed summative *effects* which were not absolute, but indicated some sound-meaning pattern. He argued that saliency (of sounds or words) will always be within a given context, such as a poem, and not relative to their frequency over a whole language (although I would suggets that we may have an unconscious sense of what constitutes normal phoneme- or word-frequency, which can be activated by the right stimulus and thus used to register the relative unusualness of an item: clearly, something like this can operate for unusual words). Perhaps Hymes' most important observation was that Lynch's approach was *partial*: the tonal organization of a poem may operate above and below the phoneme level, and that it is necessary to look at distinctive features: "There is no reason to exclude from experimental study any dimension along which sounds can be grouped" (Hymes 1960: 130).

In 1962, Maxime Chastaing published an article summarizing no less than thirteen formal and one informal experiment investigating the linkage between attributions of brightness and darkness and vowel height and backness. He started with seven minimal pair experiments; the first asked subjects to judge 5 vowels ([a], [e], [i], [ɔ] and [u]) in terms

---

[33] For a detailed description of her methodology of weighting of vowels in a poem, see Macdermott (1946) and for a description of syllable-weighting, see Priestly (1994).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of their brightness or darkness, while the rest required them to change the vowels in words (naturally-occurring French words and artificial monosyllables) to make them darker or brighter. Subjects were grade-school, high-school or college students; their numbers ranged from between 20 to 168. In every case, [i] or [e] was chosen by a significant majority to represent brightness and [u]or [ɔ] darkness. Chastaing's conclusions were that "there are such quantitative differences or similarities between the ranked vowels that certain front vowels should be considered, qualitatively and absolutely as bright and the back vowels as dark," and, quoting Gombrich and recalling Jakobson (1960), quoted above: "When we say that *i* is brighter than *u*, we find a surprising degree of general concert" (Chastaing 1962: 4).[34]

The latter part of the article describes his search for lexical proof that there exist statistically significant correlations between vowel types and meanings of darkness or brightness. More specifically, he was looking for minimal or near-minimal pairs of the *gleam-gloom* type. However, this was assuming firstly that surrounding consonants play no determining part in the brightness symbolism of words,[35] and secondly that conventional referential language will be very significantly patterned by phonetic symbolism.[36] Baffled by the intransigence of the French lexicon in matters of luminosity (though apparently not in matters of size), he declared, "If it is easy to collect oppositions

---

[34] E.Gombrich 1960 *Art and illusion* (p370) Panthcon Books

[35] This should bc sct against Marchand's (1959: 150) rcmark on thc phonosymbolic influcncc of codas, quoted above.

[36] But scc Wcrncr and Kaplan (1963) for an authoritativc rcfutation of this point of vicw.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(*apophonies*) in which sizes are contrasted... it is difficult to find any in which light contrasts with darkness. This fact poses a problem which we offer you without being able to resolve it: *why do people, who seem to us today as predisposed towards the symbolism of bright and dark as to that of large and small, practice the latter verbally more than the former*?" (Chastaing, 1962: 7, emphasis added). Perhaps the question itself was in error.

Chastaing was searching for a word count which would statistically support his subjects' very strong intuitions about vowel types and luminosity, and he explored the lexica of three languages: French, Italian and English, drawing his brightness/darkness samples from dictionaries of synonyms, as did Newman (1933). Apparently he found what he needed, although admitting that his statistics might be in error and done too hurriedly (Chastaing 1962: 10). But his investigations seemed to show tendencies or drifts rather than clear instances of symbolism (in the French case, out of 57 'bright' words, 36 had front vowels and out of 32 'dark' words 19 had back vowels). This is exactly as one would expect, given that there are other historical shaping pressures on the lexicon. Moreover, without comparing the above figures with those of normal distributions of vowels, and without also examining similar figures of vowel distributions across other polarities (e.g., words for strength-weakness, angularity-roundedness), this part of his study remains suggestive but hardly convincing.

Yet another attempt to understand the relation between phonetic symbolism and the lexicon was provided by Bolinger (1965, 1992). His special contribution was to emphasize the force of secondary phonetic symbolism through describing sound-meaning clusterings

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(phonosymbolic paradigm formation) which he pursued with a poet's hypersensitivity to both sounds and lexical associations: "*Sarcastic* is a cutting word in more than its etymology; part of its tartness comes from the insinuation of *sour* and *caustic*" (Bolinger 1965: 201). The introduction of association as a powerful but hard-to-delimit influence in the structuring of the lexicon strongly recalls de Saussure, who describes associative relations in language as those "not supported by linearity. Their seat is in the brain; they are part of the inner storehouse that makes up the language of each speaker.... Mental association creates other groups besides those based on the comparing of terms that have something in common; through its grasp of the nature of the relations that bind the terms together, the mind creates as many associative series as there are diverse relations... There is at times a double similarity of meaning and form, at times similarity only of form or meaning. A word can always evoke everything that can be associated with it in one way or another... *A particular word is like the center of a constellation*[37]*; it is the point of convergence of an indefinite number of co-ordinated terms*" (de Saussure 1959: 123—126, emphasis added)

Bolinger observed in the introduction to the first of two important essays on the topic, Word Affinities, that there is "a fundamental principle of the lexicon: the tendency of a form to gather about it an aura of meaning which clings to it in spite of the more strident demands of practical communication. It seems to me that verbal counters have both an x value and a picture value, and that the two are often in conflict. For the transactions of language, the x value dominates; for the allusions of language the picture value

[37] Bolinger actually uses the term "constellation" for these kinds of clusters (Bolinger 1965: 198).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

dominates.[38] The x value is potentially variable... the picture value is absolute-- it decrees that (a word) once embedded in the language forms semantic alliances that make it impossible to dislodge (except by a slow historical process), and once embedded in the mind of the speaker, becomes more than a substitute for a thing-- to a certain degree it *is* the thing, or part of it" (Bolinger 1965: 191).

According to Bolinger, even Jespersen did not go far enough in admitting the fullest implications of sound-meaning connections. What Bolinger emphasized was the "cumulative tendency of symbolism within symbolism-- a symbol may symbolize not a thing but another symbol; most often, however, it suggests both the thing and other symbols. A great deal of sound suggestiveness, as well as no small amount of philosophy, has resulted from symbols which stand at just such a second or third remove from things" (Bolinger 1965: 192-193). He described the lexical clusterings resulting from sound-sense associations as follows: "Once a partnership is fixed between a sound and an idea, nothing more is needed to cause that sound to give the cast of its idea to many words, alike in sound, that previously symbolized something different. If in addition there is some resemblance between the ideas themselves, eventually an entire family—by adoption—may knit itself together" (Bolinger 1965: 194).

Like Marchand (1959, 1969), for phonosymbolic purposes, Bolinger analyzed a word into onset (alliteration) and rime: "These two factors interlock with sound symbolism; but as

[38] This distiction more or less exactly parallells Sapir's, (1927) of the referential and expressive dimensions of language.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

they frequently serve to support a symbolism that is weakened or lost, they deserve to be separately entered" (Bolinger 1965: 199). In concluding the first essay, he cautioned that "It is tempting to imagine that all words we use are in some way contaminated with a suggestiveness that points beyond their literal meaning,[39] as more of them than we are wont to realize do, in fact, so point, but the argument can be overemphasized" (Bolinger 1965: 201).

In a second essay, Rime, Assonance and Morpheme Analysis, Bolinger explores the fluidity and range of sound-meaning constellations when etymological constraints on relationship are loosened.[40] Thus, after first observing that the all-or-nothing viewpoint will not serve to describe word-clusterings in this kind of study—the phonological principle of identity versus difference musts be replaced by one of "graded similarity versus dissimilarity"—Bolinger lays out five analytic principles: "(1) etymology is irrelevant... (2) given phonemic...and semantic similarity of two forms, we have enough to justify analysis into two separate morphemes (i.e. onset and rime) one or more of which have constant meaning and are held in common, the difference being due to other morphemes not held in common and having meanings (including zero meaning) of their own... (3) syntactic classes may be crossed (*sing* is a verb, *song* is a noun) (4) that a distribution as narrow as two forms is sufficient to establish a morpheme (5) that

[39] If Werner and Kaplan (1963) are correct, then it is not just tempting but accurate because there is a core physiognomized infant word at the center of each 'unmotivated' adult lexical item which is the essential generator of meaning.

[40] Amongst others, Smith (1912: 51) and Jespersen (1933: 285) have remarked that the formation of these kinds of word-nexi is a synchronic, not a diachronic, issue. For a confusion on just this point, see Brown (1994) who excludes from her constellations all words less than 100 years old, on the grounds that new coinings cannot be expected to be phonosymbolically influential, although common sense would suggest

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

difference in phonemic shape may be rationalized (*song* is an alternant of *sing*)" (Bolinger 1965: 206).

Bolinger was aware that to carry these descriptive principles to their ultimate logical conclusion could result in absurdities. To clarify, he gave "a few absurd examples" in which the rime /ɜrk/ meaning "energy" is prefixed by various 'morphologically meaningful' consonants, for example "*w/ork* 'apply energy' (prefix also in *will*)" (Bolinger 1965: 210).[41] After a discussion of additional principles needed to remove the lunatic fringe from such a study, Bolinger arrived at an important distinction between "neutral morphemes," and "affective morphemes" or "phonesthemes."[42] He observed that membership of one or the other groups (like that of constellation membership) was not an either-or affair, but a question of location on a continuum: "It would be impossible to say where the neutral ends and the affective begins. I prefer to believe that once a phonation and a meaning are attached, the two thenceforth are felt to be appropriate to each other and become potential centers of phonesthetic radiation" (Bolinger 1965: 218).

---

that the reverse is more likely to be true.

[41] But even in this apparently silly instance, if one clusters together such words as *wish*, *will*, *willful*, *would*, *want*, *won't*, *whim*, *whimsy*, *wayward*, *wham*, *wallop*, *whack*, *whip*, and so on, one can see a less absurd justification for *w-* as an independent prefix with a meaning something like energetic self-centeredness.

[42] Bolinger (1992: 28) described the phonesthemic usefully and simply as follows: "Phonesthemes differ from onomatopoeia in being vaguely analyzable, and from morphemes in not being PRECISELY analyzable (such as the *un-*, *-test-* and *-ed* of untested). Again, we must note the principle of more-or-less, rather than either-or, in the description.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Moreover, he recognized that the phonesthemic system is very extensive, making for great numbers of interrelated phonosymbolic word-clusters: "It is generally recognized that English contains a pool of forms interrelated through rime and assonance. *What is not appreciated is the vastness of the pool.* A true description of this part of English shows not here and there a few pristine forms adhering together and forming neat little bundles of morphemes, but rather shows form A merging with form B, B with C (and farther away from A), C with D (and still farther away from A) and so on, until all resemblance of A with remoter steps is lost." (Bolinger 1965: 219, emphasis added). He distinguished between full morphemes (i.e. phonesthemes) and "sub-morpheme differentials," elements such as the diminutive /i/ (as used in *movie*, *lovey-dovey*, *comfy*, *taxi* and *buggy*) which are not easy to classify, stating that "if we can show enough regularity in use," that rimes and assonances can be classified as morphemes and not "mere differentials" (Bolinger 1965: 220).

In this essay, the standard morpheme definitions were widened to take in the phenomena of expressive constellations, but the author gave many cautions. Although "within a comparatively narrow zone, formulation is both possible and valuable, in a wider zone it is possible but valueless ...a formulation that takes care of an insignificant number of forms is fruitless," while "in the remaining zone it is impossible..." This point is "borne out by the array of forms—especially phonesthemes—which are too fluid to be penned within limits or which, when one attempts to limit them, lend themselves to contradictory formulations... The sensible answer is to recognize that there are units which show two-way (or three-way or multifarious) resemblances, to describe the latter fully, but to accept

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the units as organic entities... From long practice in matching and cultivating morphemes for associative effects, the poet, who knows instinctively that *rasp* is a rougher word than *file*, can give no little to the linguist where phonesthemes are concerned" (Bolinger 1965: 225-226). This last remark clearly shows the author's understanding of poetry as a place where phonetic symbolism may be found and studied in particularly high concentration.

In a later article on Sound Symbolism (1992), Bolinger added the concept of "*postural*" sound symbolism, meaning "the slight or heavy distortions of the sound wave that accompany movements of the body, especially facial muscles. The hearer automatically refers these changes to the gestures that cause them, and a speaker is 'heard' to smile, frown etc... This in turn may be reflected in certain of the phonesthemes: an initial /v/ is produced with a sort of snarling expression, helping to create the constellation *vicious*, *venomous*, *vile*, *vindictive*, *vituperative*, *vitriolic*, *viperish*, *vixenish*, often spoken with exaggerated emphasis on the /v/" (Bolinger 1992: 29).

He also further developed the idea of secondary phonetic symbolism with the term "secondary phonestheme." "By more or less accidental association, parts of words come to carry a significance of their own, often reflecting the meaning of some dominant word carrying that part... Constellations of phonesthemes infect unrelated forms with hints of their meaning: a *swan* moves with wide, sweeping strokes reflecting the *sw*-words" (as in the constellation *swell*, *swarm*, *swerve*, *swoop*, *sway*, *swing* etc.) (Bolinger 1992: 28). He noted that secondary phonetic symbolism is an important support mechanism for lexical retrieval, since it helps us to form many extra sets of paradigms, useful cross-references as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

it were: "Words wear as many symbolic guises as there are ways our memories organize and store them: big words, high-pitched words, learned words with ponderous affixes, acute words for activity, words that absorb their senses and become silly or grave" (Bolinger 1992: 29).

A very different exploration of the relation between phonetic symbolism and the lexicon was offered by Heise (1966). His basic data were evaluations of 1,000 non-function words of high frequency of occurrence, judged along the three meta-dimensions of the Semantic Differential: Evaluation, Acitivity and Potency by 342 male navy enlistees.[43] He analyzed them in terms of the occurrence of the words' 45 compositional phonemes to see if the presence of certain phonemes in these words were predictors of their scores on the Semantic Differential. Just over one-third of the phonemes seemed to correlate with extreme evaluations of the words they were found in, i.e. the phonemes occurred in a sizeable enough number of words with scores significantly deviating from normal.

Heise then compared, using three different statistical methods,[44] mean evaluation scores for the individual phonemes with the overall evaluation scores for the words. He found a very modest correlation (of between 5% and 6%) indicating that phonemic composition would give a one-in-twenty chance of semantically foretelling word evaluation: a predictor, but a weak predictor.[45]

---

[43] Although Heise does not explicitly say so, these were presumably derived from scores on scales with high factor loadings on the metadimensions such as good/bad, fast/slow, hard/soft and so on.

[44] All three produced numerically very similar results.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Next, Heise compared his results with those of Miron (1961), who had obtained Semantic Differential evaluations for 10 phonemes, using artificial CVC stimuli. Ranking Miron's and his own scores and comparing them, 7 out of the 10 phonemes showed very different ratings on one dimension, and 3 out of that number on more than one. Heise's own comment on the most obvious discrepancy is striking: "On the Potency dimension, Miron's measurements place /g/ in the first rank position (i.e. it is hard); in the present study it is tenth (i.e., soft)" (Heise 1966: 22).[46]

In the discussion of his results, Heise suggested various explanations for the discrepancies between evaluations gleaned from his data and those of Miron's. His most interesting suggestion is that "A... possible way of reviewing the conflicting results... rests on the hypothesis that *a phoneme may be associated with multiple meanings* in the same way that words (morphemes) are sometimes homonymous" (Heise 1966: 25, emphasis added). I would suggest, following Grammont (1933) and Werner and Kaplan (1963) that a better analogy would be between the same word used literally and metaphorically. Phonemes may indeed be associated with more than one meaning, but all derived (i.e. non-auditory, non-onomatopoeic) meanings could still relate back to a core imitative and onomatopoeic

---

[45] I am indebted to Dr. Diane Bradley for this remark and for her help in understanding Heise's tables.

[46] This is a detail, a significant one. Intuitively, one feels that Miron's results have more to say about the actual phonetic symbolism of /g/ than Heise's. Grammont (who does not specifically mention individual stops) remarks: "Occlusives, striking the air with a sharp blow, contribute to the expression of a sharp sound whose timbre is indicated by the vowels" (Grammont 1946: 387). Marchand, discussing instances of /g/ in word coining, says, "It occurs (alone) almost only with words denoting guttural sounds or such as resemble guttural sounds. It is therefore chiefly found before velar vowels... /gl/ is an initial with words expressive of the idea 'light shine' as in *glass, gleam*... /gr/ is probably of sound-imitative origin with words denoting deep-toned, grumbling, inimical or menacing noises (Marchand 1959: 166-167). Most or

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

meaning. This central meaning will allow of many 'translations' and extensions, and perhaps given the right expressive contextual need, complete semantic reversals, as in e.g., "That was a really *cool* hot bath I just took."[47]

Heise's original method was to extrapolate phoneme evaluations from word evaluations without allowing for the effects of other phonemes in the words judged (let alone the influence of the *meaning* of the word, which was presumably what the original subjects thought they were evaluating, unless some of them were phoneticians besides being navy enlistees). Miron's artificial CVC words were all minimal pairs, carefully designed to contrast methodically 5 consonants in both initial and final positions and also to contrast the 5 medial vowels. (Miron 1961: 624.) As Heise himself put it, "In studies employing artificial words, co-occurrence of phonemes is a controlled factor; in this study, (i.e. his own) the factor was not controlled" (Heise 1966: 25).

The connection between markedness in syllable structure and expressiveness was examined by Fudge (1970). He proposed 7 categories of expressive words (onomatopoeic, movement-related, affective, hypocoristic, nursery, colloquial and abbreviated) of which the first three are "the most important and numerous" (Fudge 1970: 162—163). He stressed, following Bolinger, that measuring expressiveness must be in

---

all of these meanings described are clearly closer to "hard" than to "soft."

[47] Just as, lexically, there can be metaphorical meaning-reversals (as in "You lookin' real *bad* tonight," meaning "You look very good"). But these are surely marked, and based on a particular linguistically subversive effect which still assumes the primary meaning of bad as the opposite of good. For reversal of standard /i/-brightness symbolism, see Fischer-Jorgensen (1978: 84), and for other reversed or unexpected phonosymbolic effects, see Diffloth (1994) and Priestly (1994).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

terms of "a continuum, rather than a series of discrete steps" (Fudge 1970: 163), that expressiveness can be context-dependent and that it can vary from speaker to speaker.[48] Although he quoted Bernard-Thierry's four-tiered taxonomy of expressive words in which the groupings "can be regarded as successively more abstract stages of figurative application of onomatopoeic words" (Fudge 1970: 164),[49] he did not attempt to use the Grammontian insight which was implied. Instead, he attempted to fit expressive words chosen from an array of languages (such as Bantu, Malgache, Finnish, Hokkien and Javanese) into his own categorizations.

In the central part of his paper, Fudge set out to demonstrate the possible linkage between "the semantic feature 'expressive'" (Fudge 1970: 166) and phonological features of syllable types in English. Out of all possible syllable types, he calculated firstly the *usage* (what proportion of a syllable-type in question was actually used in words) and then their *restriction* (what proportion of actually-used syllables were restricted to expressive words). He found a significant correlation between the presence of an initial consonant cluster, or initial and final clusters and restriction, and also between restriction and syllables beginning and ending with the same sounds (palindromes).[50]

---

[48] For example, his criteria of expressiveness in the syllables he examined were based entirely on his own idiolect, and he stressed that "no claim can be made for the present study beyond that of representing my own judgements" (Fudge 1970: 164).

[49] S. Bernard-Thierry 1960 Les onomatopoees en malgache *BSL* 55 240—269. Inexplicably, Fudge excludes Bernard-Thierry's fourth group of expressives (soundless onomatopes) from this description.

[50] In attempting to explain this phenomenon, he quotes Hymes (1960) on methods of giving prominence to sounds in poetry, amongst which is *repetition*, including that found in the palindromic syllable-type mentioned above, and remarks that Hymes' observations imply that "*the principle is valid for the language as a whole rather than just for poetry*" (Fudge 1970: 172, emphasis added). Since in this section I am tracing the argument that poetry makes use of phonetic symbolism in especially prominent and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In looking at palindromes and the relation between them and expressiveness, Fudge suggested two lines of inquiry: either these forms are restricted because they carry a meaning of childishness and "(1) Childishness is the root form of 'expressiveness'[51]: onomatopoeias, movement verbs etc. are concrete and specific, in contrast to the more abstract and general terms that only the mature can handle" or "(2) There is a universal tendency for certain phonological features, notably repetition, to correlate with certain semantic features on the grounds of phonetic symbolism... The fact that repetition is typical of young children's speech is quite unrelated to this—thus, children's words and nursery words are only, so to speak, associate members of the category of 'expressive' words... As far as English is concerned (2) appears to be supported rather than (1) since 'expressiveness' also correlates with complexity of syllable structure... which is certainly not typical of young childrens' speech." (Fudge 1970: 178).

Nothing, however, says that what is a feature of childish speech is not being carried into adult speech in a more elaborate *form* than a child could manage, but still following a basic child-speech strategy. Consonant clusters (as either onsets or codas) are only one feature associated with expressiveness; it is true that these are difficult for young children to articulate. It is also very likely that acquisition of secondary phonetic symbolism via phonesthemic constellations of the kinds described by Bolinger is probably relatively late in language development. Yet, to take the most obvious example, vowel features such as

---

studyable ways, providing in concentrated form what is found quite normally in language at large, I include this remark as one more tiny piece of evidence.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

anteriority have been correlated with expressiveness by many authors quoted in this and the preceding chapter, and these associations are remarked on as being as, or even more, common to children as to adults (see Jespersen 1933, Reichard et al. 1949).

Fudge also noted that there was a relatively lower incidence of syllabic expressiveness associated with the presence of alveolar consonants: "Alveolar is the most frequent place of articulation, and values of R (i.e. restriction) are notably lower for this place than for any other" (Fudge 1970: 171), concluding "From many points of view, Alveolar is the unmarked[52] place of articulation" (Fudge 1970: 177). This distributional gap was addressed *in absentia* by Wescott (1971a) through an examination of the features of labiality and velarity and their connection with one particular form of expressiveness: derogation. The range of derogating terms treated was wide, including names for "tabooed objects and actions" together with "names for racial, national and ideological groups that are targets of xenophobia. Moreover, derogaratory velarity and labiality appear not to be confined to stops but rather to extend to fricatives, nasal and glides and even to low and back vowels... It should be added that... in producing derogatory connotations, velarity is more efficacious than labiality, and consonantality more efficacious than vocalicity" (Wescott 1971a: 123).

Wescott treated seven semantic domains, comprisinging "ethnic, sexual, extrusive, ingestive, psychographic, malefactive and miscellaneous" (Wescott 1971a: 124).

---

[51] Again, see Werner and Kaplan (1963) for a convincing argument that this is indeed the case.

[52] This would render it therefore least available for expressive uses of the three major oral areas examined.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Categories were further subdivided, e.g. "The extrusive category can be divided into the following subcategories: defacating (dump), flatulence (fart), urination (leak), vomiting, (barf), belching (burp), weeping (bawl), and miscellaneous (pimple)" (Wescott 1971a: 125). All his examples, he claimed, were characterized by "phonosemy, or correlation between sound and sense, to a degree which exceeds that occurring in most semantic domains" (Wescott 1971a: 125).

In his description of the structural makeup of derogatory base-morphemes, he departed radically from Marchand (1959, 1969) and Bolinger (1965) in two ways: firstly, he analyzed them as having a tripartite (rather than dual, onset-rime) structure, "coda, onset, nucleus," and secondly, he claimed that it is the coda which expressively dominates: "the most crucial phonic factor is the final consonant or consonant-cluster; the factor of intermediate importance is the initial non-vowel or consonant-cluster; and the least crucial factor is the stressed vowel" (Wescott 1971a: 125).[53]

His evidence for the "maximal importance of codas and the subsidiary importance of onsets" was "the formal roles of rhyme and alliteration, respectively in English poetry and poetic discourse (this latter term covering such items as song texts, literary titles and commercial slogans)." (Wescott 1971a: 126).[54] However, the author then rather startlingly

---

[53] He introduced the term "pentestheme": "though the term is the writer's, the concept is perhaps a century old..." This was a base morpheme of very specific segmental composition, "obstruent, plus-continuant-plus-vowel-plus-continuant-plus-stop... held to be more likely than other forms to convey special sensory or emotional effects" (Wescott 1971a: 126, fn).

[54] This has some plausibility, but since Anglo Saxon poetry relied largely on alliteration to bind together the pre- and post-caesura portions of its lines, there is historical evidence for giving more prominence to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

observed, "The probable reason for the minimal importance of syllabic nuclei is the relative frequency of vowel gradation in grammatical sequences like *sing*, *sang*, *sung*, or nongrammatical sequences like *zig-zag*, which accustom English-speakers to the notion that *vowels are mutually substitutable*" (Wescott 1971a: 126, emphasis added). In other words, the same minimal pair or triad phenomena that had been used to support claims of pervasive vowel symbolisms of size or distance (see Chastaing 1958; Orr 1944; Thorndike 1945) were here surprisingly employed to attempt to show *that vowels make no symbolic contribution to the expressiveness of a word.*

Like Fudge (1970), Wescott noted that "homophonation" i.e. total or partial (as in *konk* or *blab*) palindromy, played an important role in expressiveness. Getting around the problem that onsets do have a well-attested symbolic force (Bolinger 1965; Marchand 1959, 1969), he also claimed that there are *non-iconic* sounds with derogatory impact: "symbolic signals—that is, forms which are arbitrary in composition and hence restricted to particular languages or genealogical language groups" such as the onset clusters *sn-* and *sl-* (Wescott 1971a: 129). It is therefore possible to combine "derogatory symbols... with derogatory icons, such as labial or velar articulatory features. Examples are *snag* and *slime*" (Wescott 1971a: 130).

Wescott outlined four aspects in applying and interpreting the principle of labiovelar derogation: phonetic, phonotactic, morphophonemic and glossogenetic, meaning more and

---

onsets. Wescott himself (1987: 68) made a point about the Middle English changeover in poetic technique as having conditioned the phonosymbolic importance of rimes.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

more extended groupings of "labials" and "velars" on grounds of associations.[55] But of course, such extensions of sound-class memberships may lead to "mutually contradictory affiliations" (Wescott 1971a: 132). He also listed six "primitive articulation" speech traits productive of negative connotations, including monosyllabicity, syllabic repetition, cluster-avoidance and palindromy. However, it could be argued that these are all primarily intensifying strategies (as in *teeny-weeny*, *pop*, and so on) and carry no intrinsic charge of negativity. Like Fudge, he concluded by emphasizing the impressionistic nature of his survey, remarking, "Ideally what is needed is an exhaustive study that is not only statistical in nature but computer-aided as well" (Wescott 1971a: 136).

In another article, published the same year, Wescott (1971b) made a survey of the nature of iconism in both animal and human behavior.[56] He noted that in the animal world, "of all types of signals, those which are agonistic—that is, aggressive or threatening—seem most iconic. The evidence for this is the behavioral fact that the mating signals of a given animal are normally responded to only by conspecifics, but its threats are heeded, and generally avoided, by animals of many other species. The reason for this interspecific response is probably the fact that threat-signals (such as strident sounds, violent gestures or repulsive odors) so assault the distal senses as inevitably to suggest, and thereby to render unnecessary, an actual tactile assault" (Wescott 1971b: 418). Clearly here, *iconic* was

---

[55] For example, two sounds might participate in an alternation (as in the case of the initial /g/ of golden and initial /y/ of yellow's representing the morphophoneme {G}) or three sounds be developmentally associated (as when a child goes from the 'genophone' {W} to the three separate phones /r/ /l/ and /w/) (Wescott 1971a: 131).

[56] Yet significantly, Darwin's *The expression of the emotions in man and the animals* was not cited in Wescott's bibliography.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

being equated with *universal*.

On examining the sign language of the deaf, he observed that the gestures of American Sign Language are "more or less stylized versions of originally iconic signs" (Wescott 1971b: 418), and strongly suggested that this was true of writing systems as well. "There are several kinds of evidence... that writing has, on the whole, evolved from a predominantly iconic to a predominantly symbolic state" (Wescott 1971b: 420).

After his pronouncements (Wescott 1971a) on the replaceability and lack of stable derogatory symbolism to be found in vowels (see Householder (1946) for an opposite view of /ʌ/) it comes as a shock to see this author now admit vocalic symbolism as a widespread force in language, detailing "the three chief axes of (vowel) polarity" as "front vs. back, high vs. low and spread vs. round"[57] In surveying consonants, he pointed out "all phonic overlap encourages synonymy" citing the low incidence of palindromy in all languages and the tendency for palindromes to show "extensive synonymy": they "share a phonic device," and "rather than imitating each other... imitate themselves and therein resemble each other. English examples are *tit* = *boob* and *cack* = *poop*" (Wescott 1971b: 424). He claimed that there is also morphemic iconism; for example, marked forms like *larger*, *largest*, are longer than unmarked forms (*large*). Moreover, "iconism in syntax is, if anything, even more conspicuous than in morphology. In most of the world's languages,

---

[57] Wescott seems to be the first to describe this last grouping in just this way. In English it opposes the back rounded vowels to /iy/ /ey/ and /ay/ (one wonders how he would treat /ɔy/, since it belongs to both classes). His point is convincingly illustrated by "*yay* 'a cheer to spread encouragement among one's fellows' vs. *boo* 'a jeer to circumscribe the hopes of one's antagonists'" (Wescott 1971b: 420).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the normal declarative word order in sentences is subject-verb-object (Greenberg 1966: 76 ff.).[58] And this word order represents, better than any other, the actual order of transitive events." (Wescott 1971b: 424-425).

Claiming an even wider use of iconicity, he stated that "many of the over-all characteristics of languages seem to mirror the non-linguistic world" (Wescott 1971b: 425), giving numerous comparisons: for example between the phonemic composition of morphemes and the composition of kin-groups by individuals, or between linguistic and social hierarchies or between linguistic asymmetries and human behavioral and anatomical asymmetries. He came to a number of conclusions, including the continuous[59] (rather than discrete) nature of iconicity, and suggested three possible reasons for the "extensive co-existence of icons and symbols in language."

The first saw language as originally exclusively iconic but in the gradual process of becoming exlusively symbolic with "symbols replacing icons at a constant rate." The second hypothesis is that "when promordial icons become symbolic" due to processes of linguistic change, "new icons are introduced at such a rate as to maintain the proportion of the two sign-types at a roughly constant level." The third is that "the relative proportions of iconism and symbolism in language fluctuate slowly but rhythmically—somewhat in the

---

[58] Greenberg, J.H. (ed) 1966 *Universals of language*, 2nd ed Cambridge, Mass. MIT Press

[59] This seems to be the general view on both iconicity and phonetic symbolism: that they are best understood as continua. But for an interesting distinction between universal and local symbolism and a restriction on the continuum model, see Priestly (1994).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

manner of representational and non-representational art-styles." (Wescott 1971b: 426—427). In a sense, the second and third hypotheses are static and dynamic versions of each other; all three offer testable accounts of the relations between the two types of signs, but the second and third seem intuitively more plausible.

One aspect of linguistic iconism in a group of non Indo-European languages was reported on by Nichols (1971). She surveyed a large number of aboriginal languages of northwestern America, noting that they display expressive consonant shifting to symbolize diminutiveness. Two phonic strategies are used; one is increased "hardness"[60] of articulation (lenis to fortis, continuant to noncontinuant, nonglottalized to glottalized). Another is raising of pitch, achieved either by fronting palatal consonants so that plosive bursts or fricative noise increase in frequency, or by moving certain alveolar consonants to palatal positions, thereby raising the locus of F_2, which is apparently critical for perceiving consonantal high frequency (Nichols 1971: 830).

Nichols remarked that, given a group of antonymic attributes, "extensions of the semantic domain, such as bright-dark, light-heavy, quick-slow, near-far, and... endearing and pejorative" (Nichols 1971: 828), diminutiveness could be associated with a predictable clustering of polar choices. She remarked that although pitch-raising is generally associated with smallness, fortition is an unusual phonic correlative: "Only Korean has been found to make systematic use purely of hardness... and note that the Korean shift is not strictly diminutive" (Nichols 1971: 840). She cautioned that the semantic clustering

[60] Perhaps more properly described as hardening or fortition, since it is a process.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and apparent equivalences of "concepts such as light, bright, quick, cold (as opposed to dark, heavy, slow, warm)" needed precise definition "as a necessary preliminary to examination of symbolic properties" Moreover, any form of phonic expressive shifting "cannot be studied in isolation" but must involve study of other shift types in the language (Nichols 1971: 841). Her conclusion was that diminutive symbolism must be "an areal feature spread by borrowing" (Nichols 1971: 839). If she was correct in this, then it may be that other widespread phonosymbolic phenomena need not all be explained in terms of absolute, intrinsic universality but, depending on geography, sometimes as phonic strategies borrowed from linguistic neigbors.[61]

In the same year, Tanz (1971) on a different quest presented "further evidence for a universal sound symbolism... drawn from natural languages rather than experience" (Tanz 267: 1971). Observing that a proximal-distal contrast is often reflected in the contrast of /i/ and /a/, she surveyed terms for "here" and "there" in 43 languages.[62] From her list of forms, she found six linguistically unrelated minimal pairs, all of which showed an association of distal members with /a/ and proximals with /i/. Moreover, another six languages showed front-back vowel alternations plus other syllabic changes (for example, the Aztec *nika-n*: 'here,' *onka-n*: 'there'). She found that three more languages took the word for "here" as the root, adding a syllable to it to construct "there": "in all cases, the

---

[61] This would put a new slant on the observations of Jesperseu (1922, 1933) and on those of Chastaing (1958, 1962) who draws so many of her examples from both French and English.

[62] However, these comprised only 18 language families: 11 Indo European, 4 Dravidian, 3 Ural-Altaic and 3 Semitic, so the spread was not as complete as could be wished.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

vowel in the syllable added is /a/" (Tanz 267: 1971). She also mentioned (though unfortunately without specifying, since she wished to concentrate only on vowel contrasts) minimal pair cases where the alternation was consonantal. Thus, she found phonosymbolic correlation in around 33% of cases, but this figure is over-conservative since she suppressed the consonant-driven examples.

Tanz then turned her attention to irregular English verbs which frequently display a contrast between /iy/ or /ɪ/ in the present and another vowel in the past, arguing that the association of a high front vowel with the present tense shows chronological proximity. She found that around 25% of verbs had this vowel pattern, whereas only 3% had /ɪ/ in the past and not in the present. Finally, she briefly considered distance and proximity in relation to pronouns, and noted that the Japanese intimate *kimi* and formal *anata*, (again associating /i/ with the proximal form and /a/ with the distal), could act as phonosymbolic reinforcers: "If the /i/-/a/ proximate-distant sound symbolism does operate in Japanese, the use of *anata* and *kimi* would have an extra reverberation. In addition to their reference and their "sense" as determined by the set of sociolinguistic rules governing their usage, the words would have a mimetic halo. Each utterance of *anata* or *kimi* would reinforce a sense of distance or intimacy at a level of communication which might be called the "poetics" of social interaction" (Tanz 270: 1971).

This is an important point: if phonetic symbolism is operating in the Japanese case, then Tanz' example may well be a representative instance of how the mechanism works. The

---

Semitic, so the spread was not as complete as could be wished.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

use of core, high-frequency morphemes (common 'information' or 'content' words such as *nice, girl, fix, get* and so on, but also function morphemes—pronouns and prepositions, deictics and certain affixes like *un-* or *-ing*) could serve to strengthen a tendency that was already there—or to instill, through the practice of constant sound-meaning association, a language-specific phonetic symbolism.

Tanz, arguing for the presence of phonetic symbolism in language, was looking for a modest piece of systematicity. On the other hand, Gamkrelidze (1974), in a theoretical paper on the arbitrariness of the sign, was concerned to limit the scope of phonetic symbolism to what he termed "horizontal relations" (Gamkrelidze 1974: 102). He defended the original concept of de Saussure, but with an interesting concession, claiming that the relation of signs to each other, as well as to their referents, was extremely important. *Vertical relations* obtain between a word and its referent, while *horizontal relations* can occur between a group of related referents or the group of words that designate them. The "opposite thesis" to Saussurian arbitrariness involves these horizontal relations.

Gamkrelidze dismissed in absolute terms the search to find strong lexical support for phonetic symbolism: "Attempts at discovering in natural languages certain correspondances between sound and meaning, between phonetic symbolism and the phonemic form of words, do not in general yield positive results. It is possible to establish, under experimental conditions, definite universal patterns of correspondances

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of sounds to meanings; *but these sounds are not distributed in the concrete words of natural languages according to the symbolic meanings expressed by them in isolation under experimental conditions*" (Gamkrelidze 1974: 105, emphasis added). This is clearly too all-or-nothing; a statement. Amongst other phenomena, it ignores the issue of expressives in languages such as Japanese (Hamano, 1986) and Korean (Martin, 1962) and the word-constellations described by Bolinger (1965).

However, he also echoed Wescott (1971b), whom he later cited, in claiming that sets of horizontal relationships amongst referents (Nichols' diminutiveness or Tanz' proximal and distal would be examples) are indeed symbolized in linguistic structures: "We may accordingly speak of one series of relationships as being motivated by the other in the content plane of language" (Gamkrelidze 1974: 106). Amongst other examples, he gave that of the (surely iconic) character of the plural morpheme: "There are languages in which the plural forms are distinguished from the corresponding singular by an additional morpheme, whereas there is no language with a reversed correlation, i.e. with a specific morpheme for singular vs. absence of an affix in the plural forms... This correlation in phonemic length between... shorter singular vs. longer plural forms is a specific manifestation of the relationship of quantity.... on the plane of signata" (Gamkrelidze 1974: 107).

Yet he refused to label such relations as iconic: instead, they were described as "in principle conventional" (Gamkrelidze 1974: 108). He denied that "relations of quantity, of the intensity of a quality, of immediate neighborhood" are imitated or translated by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

repeating that vertical (word-referent) relations are *always* arbitrary apart from out-and-out cases of onomatopoeia. Even Köhler's *maluma-takete* experiment was considered not to "involve the 'vertical' relationship; it must be referred exclusively to the 'horizontal' relations between the signs" (Gamkrelidze 1974: 108). But the point of Köhler's experiment was precisely to demonstrate that the relation between word and figure had exactly that sense of quality-for-quality correspondance or equivalence which we label iconic.

Croft (1978) took the opposite position from Gamkrelidze's on iconicity,[63] very much following Bolinger in terms of seeing it as an important, constellation-forming presence in the lexicon, and thus a great aid to lexical retrieval. Her article briefly surveyed some studies of synesthesia and iconicity and a made a strong plea for using the latter as a pedagogic resource in language teaching. She cited numerous "relational or 'horizontal' iconicities" present in language (Croft 1978: 511) such as palindromic formation or partial formation of antonyms as in the English word pair RAB-ble: BAR-on.

Given its use as a cross-referencing mechanism, the application of iconicity in the study of languages could, she claimed, be very helpful for two reasons. First, it would provide an extra 'iconic' category to label any items so described, "to 'file' things in the mind. Second, and more important, the iconic tag has a particular relevance because *it provides the student with a ready-made associative tie between the realms of sense and sound*

[63] She even quoted Kohler's *maluma-takete* experiment as an example of a "non-arbitrary relationship between kinetic and tactile impressions and sound" (Croft 1978: 512).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which other tags do not have and which plays an important part in the recall of linguistic material" (Croft 1978: 511, emphasis added).

In the same year, Fischer-Jørgensen (1978) published an account of an experiment to investigate "whether expressive words, which do not imitate sounds, have a universal basis" (Fischer-Jørgensen 1978: 80). He examined Westermann's[64] work on expressives in African languages: "New (words belonging to the category of expressives) can be formed and are immediately understood by the listeners because they are created according to definite principles. Almost all phonetic features may have an expressive value: vowel length, vowel quality, anteriority and manner of articulation and pitch" (Fischer- Jørgensen 1978: 80—81).

Westermann divided vowels into three expressive groups: 1) /e/, /ɛ/, /a/ which are front and unrounded, 2) /u/, /o/, /ɔ/, which are back rounded, and 3) /i/ which is in a category by itself: "It is a characteristic feature of these languages that the vowel /i/ does not come within the category of front unrounded vowels, but has its own specific expressive values, indicating something narrow, tight, squeezed or very dark" (Fischer-Jørgensen 1978: 81). The first two groups express very generally the meanings respectively 'intensive' versus 'extensive.'[65]

---

[64] Westermann D. 1937 *Laut und Sinn in einigen Westafrikanischen Schprachen*, Archiv f. vergl. Phonetik. This work has apparently not been translated into English.

[65] As evidence for the principle that any linguistic feature may be employed as a supplementary expressive resource through its use in a marked way, he noted that the dimension of tone, which normally combines with front vowels as high and with back vowels as low, could be separately used to add extra, unusual meaning: "For instance, a word which indicates things are normally big will normally have a low tone, but if in certain situations the thing indicated is small, the tone is changed to high" (Fischer-Jorgenson

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Fischer- Jørgensen wished to see how far the West African expressive values of vowels would be perceived "at least as latent possibilities" (Fischer-Jørgensen 1978: 82) by speakers of unrelated languages. Accordingly, he took two groups of Danish elementary phonetics students (totalling 99) and asked them to match the vowel groups 1) and 2) to 9 pairs of polar attributes taken from Westermann's description: attributes included thin/thick, flat/round, tight/loose and bright/dark. Presentation was visual; vowels were written on the blackboard in phonetic script. Fischer- Jørgensen admitted that the experimental task was forced-choice, since assigning one vowel group to one pole implied assigning the other group to its polar opposite.

Vowel judgements evoked by six out of the nine attribute pairs yielded strong subject consensus with the West African evaluations of these speech sounds.[66] Arguing against the possibility that the vowels in the words chosen to denote the attributes may have influenced the subjects' judgements, Fischer-Jørgensen stated that in only three out of the nine cases did the vowels in the epithet pairs match up with the West African values; however, in other cases there was just as high a percentage of agreement, "so this cannot be the reason" (Fischer-Jørgensen 1978: 83).

In this experiment /i/ was not included, but in earlier tests (1967), "it behaved like the other front vowels" even having "an extreme position, being the most thin, flat, bright,

---

1978: 81).

[66] The three pairs that did not were hard/soft, tight/loose and weak/strong.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

light and hard... As for the specific expressive values mentioned by Westermann, there is agreement insofar as /i/ was considered more small, tight, tense and compact than the other vowels... *but /i/ was never considered to be dark; on the contrary, it was the brightest of all vowels*" (Fischer-Jørgensen 1978: 84, emphasis added). This interesting anomaly shows possible counter-evidence against universal vocalic symbolism; moreover, the author remarked that even in (European) cases of colored hearing, "Normally... /i/ goes with the front vowels, *but there are also examples of /i/ being felt as black.* This is, e.g. the case for some French subjects and a great number of English subjects" (Fischer-Jørgensen 1978: 84, emphasis added).[67]

Fischer- Jørgensen considered the symbolic correlatives of different phonetic features, noting that though "it has generally been assumed that the front-back difference was the important one, Westermann believed that the rounded-unrounded difference was the decisive one for the symbolism" (Fischer-Jørgensen 1978: 85). He cautioned against expecting to find phonetic symbolism systematically represented in "the normal vocabulary of all languages.." and added that in investigating, "comparative or statistical methods are preferable" (Fischer-Jørgensen 1978: 85). He found the word-matching type of experiment "rather dubious because so much depends on the words chosen... since the words were at best partly expressive they (i.e. subjects in these experiments) did not know which sounds to compare" (Fischer-Jørgensen 1978: 86). He also forbiddingly commented on Grammont's idea that onomatopoeic imitation is the basis of phonosymbolic translation

---

[67] The questions here are then which vowel in the subjects' systems will be perceived in opposition as bright or white, whether choices are predictable and what kind of semantic reversal may be going on.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that it "did not permit the conclusion that (speech sounds were) really used in this way" (Fischer-Jørgensen 1978: 86).[68]

Next, the author considered briefly some explanations of phonetic symbolism: is it primarily conditioned by acoustic or articulatory sensations? He found some articulatory explanations "quite convincing..." the symbolic value for /i/ in West African languages being "'squeeze', 'press', 'narrow'... and it is probable that it is this articulatory starting point that leads to the symbolic value 'dark'... However, in many cases it is the auditory impression which is at the base of symbolic value, e.g. bright-dark" (Fischer-Jørgensen 1978: 86—87). This was followed by the larger question of why indeed call frequencies "high" and "low." Noting that deaf (Ertel, 1969)[69] and blind (Fónagy, 1963) subjects evaluated /i/ and /u/ in the same way as did the hearing and sighted,[70] Fischer-Jorgensen concluded that there are interrelations between qualities, stating that consequently in order

---

[68] But it certainly doesn't *prohibit* this conclusion; experimental verification is needed. This is where a method like Householder's (1946) survey of /ʌ/ (which would start not from a pre-determined symbolism but from the distribution of this phoneme in words, including function words) might provide crucial data. However, there are several tricky questions that would need settling-- the first being what criteria one would establish for distributions that would indicate real sound-meaning correlations for speakers. In other words, one would need to know what values would authentically constitute "significance" in claiming that a significant proportion of words both contained a given phoneme or phonemes and meant a predictable symbolism. Another question would be how many counter-examples (same phoneme, conflicting symbolism) would be needed to neutralize this conclusion, and whether indeed a set of counter-examples *would* be an adequate neutralizer. It would seem as if this kind of large scale phonosymbolic analysis, starting with the phonemes and surveying distributions of meanings, would be most easily done using the biggest available computerized lexicon with a phonemic component; this kind of dictionary must already exist for several languages.

[69] Ertel, S. 1969 Psychophonetik. Untersuchungen uber Lautsymbolik und Motivation. Gottingen

[70] I.e. "/i/ is smaller, thinner, quicker, brighter, more cheerful and kinder than /u/" (Fischer-Jørgensen 1978: 87).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to explain phonetic symbolism one must understand the general phenomenon of synesthesia.

He cited the importance of "a general cultural linguistic tradition with conventionalized synesthetic relations, e.g. the normal application of adjectives to different senses (a warm room, a warm color, warm greetings etc.) and the metaphoric use in poetry." However, he admitted that this still did not answer several questions: of the origins of such a tradition, or of striking cross-cultural similarities in synesthesias—or of why "small children have more synesthesias than grown-up people" (Fischer-Jørgensen 1978: 88).

He speculated that perhaps the phenomenon was based on generalized real-world experience (cf. the remarks on sound symbolism and phonetic symbolism at the beginning of this dissertation), or perhaps to "deeper interrelations between our senses" (Fischer-Jørgensen 1978: 88). The author cited a project of Ertel's to reduce phonetic symbolism to Osgood et al's three metadimensions of activity, potency and evaluation, but "his attempt to apply these dimensions to the symbolic use of vowels (was) not very successful." He concluded that there was "still much to be done in this field of study" (Fischer-Jørgensen 1978: 88).

"The time has surely come for a re-examination of sound symbolism" Thus Fox (1982: 55). introduced his study, using what had been considered a marginal phonosymbolic concept to help explain a problem of Semitic etymology: the presence of certain root consonants in a group of semantically-associated Semitic words. This itself was part of a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

larger problem: whether "the number of root consonants in the proto-language" (Fox 1982: 56) is two or three. Fox's explanation for the third consonant in two groups of words denoting such activities as *cutting, biting, ripping into small pieces* and so on was that it was conditioned by the presence of an initial phonestheme, related to the violent meanings of the words in question.

His argument was based on the Bolingerian idea of secondary phonetic symbolism. He also held the phonestheme to be responsible for consonant-gradation in the second word group,[71] so that "*qs- gz-* etc. may never have been separate morphemes or biconsonantal roots in Semitic, any more than the English phonestheme *sl-* ever was" (Fox 1982: 58). What is striking about this small study is the sense of phonetic symbolism's having become *respectable*, not as in the 50's and 60's debate over word-matching and universals, but in terms of one more force shaping the lexicon which must be acknowledged and explored.

Haiman (1983) examined two sometimes opposed linguistic forces which he held to be responsible for much of the shape of language: economy and iconicity. He claimed there was a linguistic dimension, "that of distance between linguistic expressions which corresponds directly to, and in a sense is motivated by, a variety of conceptual dimensions" (Haiman 1983: 781). In his terms, "linguistic distance" was not simply the distance between two elements on a string, but between a number of graded, non-

[71] Producing related forms such as *kss*: 'cut asunder', *ksh*: 'cut off', *ksr/kzr*: 'smash', *gzz*: 'cut off/down' (Fox 1982: 57).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

segmental boundaries, such as those occurring between bound morphemes, free morphemes and phrases.[72]

He gave a number of examples of conceptual distance, including causation, co-ordination, transitivity and possession and claimed each of these could be iconically reflected by linguistic distance. Thus, contrasting such sentences as *I caused the man to die* and *I killed the man*, Haiman observed that the first, analytic causative suggested magical powers, i.e. an absence of physical contact—distance—between the causer and causee. The second, synthetic causative connoted co-presence of causer and causee. He used instances from languages as diverse as Korean and Mixtec to further illustrate linguistic distance as an iconic correlative of conceptual distance.

Haiman claimed that the iconic forces in language may be opposed by "economic motivation for reduction of form" (Haiman 1983: 801), that under certain circumstances the principles of iconicity and economy might compete, and even that two kinds of iconicities might conflict Although dealing with syntax and thus not directly related to phonetic symbolism, Haiman's work is included here because, like Fox, the author was concerned to demonstrate that linguistic iconism was a respectable topic, and to claim a central place for it in the mechanisms of language.

Another explanatory use of phonetic symbolism is found in De Reuse (1986), who offered

---

[72] Thus the synthetic forms *cut* (past) or *sang* display no distance between the past morpheme and the root, but the agglutinatives *wanted* or *wished* show some distance, while the analytic forms *did wash* or *did sing* display even more.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

a phonosymbolic reason for "the occurrences of /S/ (in Santiago del Estero Quechua) which appear to provide exceptions to the regular sound laws" (De Reuse 1986: 54). The segment /S/ substitutes for the historically expected /s/ in words and suffixes with "particular emotional connotations, chiefly diminutative or augmentative... Affective or deprecatory value and not mere size is the determining factor in this process" (De Reuse 1986: 56). He found palatalization in Quechuized Spanish names, augmentative, pejorative or hypocoristic suffixes, not as a productive phenomenon, but "frozen into the morphology and lexicon almost completely" (De Reuse 1986: 61).

An article by Wescott (1987) focused on bi-morphemic analysis of expressive words, differing from his earlier (1971a) tripartite division of such words. He began by looking at Pei's[73] definition of "potentially morphemic phonesthemes," i.e. onsets, and "non-morphemic phonesthemes," i.e. rimes (Wescott 1987: 67). Onsets, such as the *cl-* of *clatter*, are considered asyllabic (in contrast to rimes) but Wescott argued in some cased that they are contracted versions of syllabic forms that constitute independent words (e.g. the *cl-* of *clatter* relates to *call*, and the *bl-* of *bleat* relates to *bellow*). These onsets are much older (deriving from Proto-Indo-European) than rimes like *-ash* or *-ump*, which are Middle English in origin. It is "probably no coincidence" that "the semi-autonomous sound-effect function" of rimes like -ash and -ump develop at the same time as rhyme becomes "an important prosodic device in poetry and song" (Wescott 1987: 68), and alliteration ceases to be the primary line-binding device.

---

[73] Pei, Mario 1966 *A glossary of linguistic terminology*, Anchor Books, Doubleday, Garden City, NY p 203

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In order to give more equal importance to rimes, Wescott proposed a new term "holestheme" (Wescott 1987: 68) for words in which both rime and onset are phonesthemic. He noted that rimes of expressive words are mostly monosyllabic, but where bisyllabic, there is often a verb suffix "*-le* or *-er* of the type traditionally called frequentative" indicating repeated or prolonged action (Wescott 1987: 68). He described holesthemes as semantically categorizable in terms of six types of reference: "in order of decreasing frequency," they are "tactile, phonic, kinesthetic, attitudinal, locomotory and visual" (Wescott 1987: 69). He also analyzed holesthemes into a root morpheme (syllabic or asyllabic) plus up to five other morphemes; he gave the example *scramble*, analyzed into six separate morphemic units with *-c(a)r-* 'to claw' (presumably Proto-Indo-European) held to constitute the root. (Wescott 1987: 70).

How phonosymbolic (in Wescott's terminology, "phonically iconic") are holesthemes? Wescott claims there are three kinds of iconism: *primary* (imitating non-linguistic reality), *secondary* (imitating other words: this is close to Bolinger's secondary phonetic symbolism) and *tertiary*: "repetition, within a word, of another sound in the same word" (Wescott 1987: 71). This last case relates to Fudge's work on marked syllable types and to Wescott's remarks on partial or total palindromy.

A hard-line behavioristic "parallel processing" explanation of phonetic symbolism was offered by Saka (1988) in an unpublished paper. According to Saka, the individual components of a word (segments) "compute by being appropriately connected to large

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

numbers of similar units" (Saka 1988: 6). The process of word recognition starts as the onset is perceived and matched with other onsets, and then as each segment arrives it activates similarly placed matching sounds in other words, until a complete segment-by-segment match is made "according to either a 'Winner-Take-All' scheme or an inhibition mechanism" (Saka 1988: 6). Saka claimed that phonetic symbolism is generated by this type of word recognition process. In the course of matching words segment by segment (both right to left and left to right), *a large number of other words are activated "which are all suppressed within a fraction of a second. But they survive long enough to reinforce the meaning of the stimulus word and make it 'seem appropriate.'"* (Saka 1988: 6—7, emphasis added). Saka argued against any intrinsic relation between speech sounds and sensory experience, except for some instances of onomatopoeia. For this author, "most sound symbolism arises because the comprehension module is constantly co-activating phonically related words" (Saka 1988: 12), and "most cases of sound symbolism can be accounted for by subregularities in the lexicon" (Saka 1988: 14).

Saka's description of the matching and activating process has a lot of plausibility; it fits well with Bolinger's concepts of sound-meaning constellations and of secondary sound symbolism. It seems a useful way of examining salient word choice of all types (perhaps especially in poetry which abounds in marked word choices) since it claims that all kinds of extra paradigms are activated in the phonetic-semantic background of a word.[74]

---

[74] For example, in the opening line of Hopkins' sonnet "I wake and feel the fell of dark, not day," it could be claimed that *fall* is doubly activated—by both *feel* and *fell*, and is phonosymbolically hovering in the background in a number of ways: phonetically, as a low back vowel (assuming lower and farther back equals darker, which fits with the poem's emotional tone and with the low vowel of *dark*), and semantically, as a reference to sin and the Fall (again, fitting the emotional tone) and phonetically again

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

However, to reduce the whole of the phenomenon of phonetic symbolism to something purely lexically self-referential, with no possible input from articulatory or acoustic sensations, is to take the language-habits explanation too far. Amongst other things, it argues that phonetic symbolism begins only when the speaker has a large enough vocabulary to cement these kinds of multiple sound-meaning associations, yet Saka himself noted that 'clang' or purely phonic (rather than syntagmatic or syntactic) associations to words were much stronger in childhood than later (Saka 1988: 9). And childhood is surely the time when the sensations of hearing and articulating will be experienced most intensely.

If Saka's views on the origin of phonetic symbolism represent the most behavioristic lexically-conditioned explanation, then a dissertation by N. Ciccotosto (1991) demonstrates the opposite point of view. For Ciccotosto, the origins of language are bound up with gesturally-generated phonosymbolic universals. It is they that iconically animate core vocabulary: "sound symbolism... should be examined carefully because of its importance to an understanding of proto-language and language origins... (It) is the logical bridge between what must have been a rudimentary and highly gestural language of Homo erectus and a more arbitrary sound-meaning language of Homo sapiens" (Ciccotosto 1991: 191).

---

emphasizing the falling nature of the vowels in most of the line (i.e. rising with *wake*, and falling in front steps with *feel* and *fell*). *Fell* also activates *hell*, the semantics being enhanced phonetically with subliminal alliteration, /h/ being a protofricative and /f/ being a fully specified fricative. And so on.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To support his thesis, Ciccotosto chose 16 basic core vocabulary items comprising firstly body parts: *neck, tooth, mouth, nose, breast*, secondly, bodily functions: *cough, eat drink, vomit, suck, swallow, spit, chew* and thirdly basic nouns: *dog, food, water*. These were translated into 50 languages from "virtually all known language phyla," making a total of 800 items. He predicted that in any set of 50 translations of a word (for example the word *dog*) "certain sub-phonemic features, such as nasal, stop, spirant, bilabial" would be found at higher or lower levels than those in the control sample, i.e. the remaining 750 words. (Ciccotosto 1991: vi-viii).

Thus, he predicted of *breast* that "nasal, bilabials and front vowels should be higher than chance/normal in the sample because they are found in the same areas most used in suckling. Since feeding is a continuous process, less than chance/normal distributions of stops should be seen"[75] (Ciccotosto 1991: 30). A number of similar feature predictions were made for every word; then the actual occurrences of those features were checked against their occurrence in the control sample.

Ciccotosto's is an interesting hypothesis, but there are two serious drawbacks to his study. The first is the often quite surprising and eccentric assignation of features to words. For example, he predicted there should be a high frequency of bilabials in the *nose* sample because they "represent the nose visually similar to the protruding possible with the lips" or in the *water* sample "front vowels should be higher than chance/normal frequency since

[75] Counter-examples (in English) are *tits, boobs, knockers, bristols,* and so on.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

they mime drinking behaviors"[76] (Ciccotosto 1991: 31). The second problem is the smallness of the sample of core vocabulary, using only 16 words, rather than, say, 50 or 100. Thus, even if he had obtained highly significant results—striking correlations of predicted features with words—all that his work would have indicated was a tendency of a small number of core vocabulary items to show possible common gestural traces. In fact, Ciccotosto's results were far from conclusive, with only one-third of his predictions borne out by fact and showing a "significant association between feature and meaning" (Ciccotosto 1991: 192).

Yet another attempt to reconcile the Saussurean concept of arbitrariness with observations of linguistic iconicity, especially phonetic symbolism, was made by Linda Waugh (1993). She pointed out that much of current literary theory is based on de Saussure's claim that the link between signifier and signified is largely unmotivated, yet in linguistics meaning and reference tend to be seen as "tightly connected to each other" since "meaning is investigated through the truth value of sentences, which is assumed to be tied to the truth or falsity of the referential situation that these sentences denote... a good deal of modern linguistic theory is based on the close interconnection of meaning and reference" (Waugh 1993: 72). Waugh was interested in pulling the two disciplines of linguistics and literary studies together, and in rethinking "how language and texts are viewed... arguing for the closeness of the tie between sound and meaning" (Waugh 1993: 72) and asserting that both disciplines had iconicity in common.

---

[76] As in the onomatopoeic French infant form: *faire glou-glou*, 'to drink,' perhaps?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Presenting a brief list of iconic structures in language (see Wescott 1971b and Haiman1983), she then asked whether iconicity also existed in the lexicon or whether it was the last stronghold of arbitrariness. Following Peirce,[77] she distinguished between three icon types: *images* (in which qualities of the form resemble those of meaning), *diagrams* (in which relations in the form correspond to relations in meaning) and *metaphors* (which she did not treat in the article).

Waugh claimed that image iconicity is exemplified by onomatopoeia (which may be partial or total, subtle as in *sniff* or overt as in *slurp*). She then cited various cases of systematic onomatopoeia, such as the association of high front vowels with smallness and brightness, and the low back vowels with opposite qualities, or the association of alveolar consonants with the diminutive suffixes, as in *brooklet* and *kitchenette*[78] (Waugh 1993: 74). However, she pointed out that the use of suffixes is in itself an example of diagrammatic iconicity. Other diagrammatic instances are to be found in word-groups with common submorphemes, such as the *tw-* of *two*, *twenty*, *twin*, *twice* and so on.[79] Phonesthemes are "characteristic of more far-reaching families of words" and may be combined with onomatopoeia. Waugh, however, cautioned that not every complex onset is phonesthemic,

---

[77] Peirce, Charles S. 1902, reprinted 1955, "Logic as semiotic: the theory of signs" in *Philosophical Writings*, Ed J. Buchler, New York: Dover 98-119

[78] This contradicts Fudge's (1970) claim about the lack of expressivity of alveolars because they are the commonest type of consonant.

[79] "This consistency of form-meaning relations across words is known as the principle of isomorphism or isomorphic iconicity, which is a specific type of diagrammatic iconicity" (Waugh 1993: 75).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

contrasting *grim*, *grisly*, *gritty*, *grotty* and *greedy* with *grin* and *gratify*. (Waugh 1993: 77—78).

Finding extensive iconicity in the lexicon, she remarked "The lexicon seems to be directed by the extremely important functional principle that formal recurrences across words are a major clue to sharing of meaning by those words. Moreover, the word-internal structure and cross-word associations thus created allow us to amass a large lexicon and to make sense of it... It is easy to see why imitation and motivation exist: these iconic relations across words act as a mnemonic; they aid us in learning, memorizing, producing, recognizing and understanding a vast number of words" (Waugh 1993: 79). Thus, "the function of sound is, first, to establish differences between words and, second, to create a myriad of form-meaning identity associations across words" (Waugh 1993: 80).

As a counterweight to seeing "excessive iconicity" (Waugh 1993: 80) in the lexicon, she looked also at the issue of constraints on iconicity. Prime is polysemy, of which homonymy "is simply the limiting case of extensive and wide-ranging polysemy" (Waugh 1993: 81). Here we find the semantic forces that pull apart sound-meaning constellations; Waugh's example is *understand*, in which a previously bimorphemic word has been lexicalized as monomorphemic, having lost its original motivation, i.e., its relation to the parts *under* and *stand*. Yet Waugh remarks that motivation and lexicalization are often partial. There exists a continuum of motivation: "while wholes are not just the sum of their parts, they are related, but not completely or in a predictable manner, to their parts" which means "that it is difficult to find an identity of meaning across words that are part of word-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

affinity relations. This would seem to deal a death-blow to the quest for iconicity..." (Waugh 1993: 82).

Her solution was to see similarity of meaning (the precondition for iconicity) not as an either/or condition but as a continuum (as have Bolinger (1965), Fudge (1970) and Wescott (1971b): "It is obvious that the majority of words in the lexicon exhibit partial iconicity. They manifest a delicate balance between full iconicity and total arbitrariness" (Waugh 1993: 83) Waugh saw the lexicon (and indeed the totality of language) as balanced between the opposing forces of iconicity, which make for "invariability, homogeneity, unity, determinacy, predictability, precision—all of which are absolutely crucial to the working of lexical meaning and language in general" and the "noniconic force" emphasizing "the pure conventionality of the tie between form and meaning, and thus a certain autonomy of each" (Waugh 1993: 84).

In an article that unconcsiously owes much to Bolinger's (1965) thinking about sound-meaning constellations (although he was not cited as a reference), Brown (1994) once again looked for systematicity in the lexicon. Using the Longman Active Study Dictionary, she surveyed constellations of all monosyllabic and bisyllabic words "beginning with a given consonant cluster or containing a similar vowel"[80] (Brown 1994: 18). She limited herself to words of 100 years or older, i.e. those that she considered had existed long enough in the language to help build phonosymbolic influence.[81] Of the six initial cluster

[80] In fact, she was looking not for similar vowels but *rimes*, as her examples made clear.

[81] Here she departed from Jespersen's (1933) and Bolinger's (1965) viewpoint: they had a strictly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

examples she gave, four showed a 50% association with a given meaning. She also included three rhyming sets in which the meaning-association was higher (60-75%).

Brown admitted that her procedures were "open to challenge" (Brown 1994: 19), since it was she alone who decided which words to include in a set, and whether or not they demonstrated the predicted semantic connections. She concluded that further, more methodical study could be very useful in this area. Suggesting that there might be applications to second language learning, she questioned whether advanced second language learners might not be sensitive to these types of sound-meaning connections.

1994 saw the publication of *Sound Symbolism*, an important anthology of writing on phonetic symbolism. The book included articles on the indigenous languages of North and South America, and on Asian, Australian, African and European languages as well as a section on the biological basis of sound symbolism. In this last section, there were essays on the association (in both animals and humans) between F_0 pitch height and depth and communication of sexual differences, agression, threat or submissiveness (Ohala, 1994; Morton 1994). Ohala's work also made a strong connection between cross-linguistic size symbolism in vowels and consonants characterized by high frequency, and a universal (i.e. animal and human) response to large and small sounds (Ohala 1994: 335—336).

---

synchronic view of these kinds of constellations, presumably since what was being represented was the lexical consciousness of a contemporary speaker. If a word has been recently coined, using the same phonestheme or phonestheme combination as a number of others of greater antiquity, this hardly invalidates its mutual influence *on* and *from* other, older words belonging to the same paradigm. Most speakers, apart from etymologists and lexicologists don't mark words for age (though certainly they may mark them for fashionability, like *awesome* or *cool* or recency of coinage, like *jiggy*).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Many of the articles, like Nichol's (1971) or Fox's (1982) claimed that phonosymbolic mechanisms are decisive in systematizing sound-meaning correspondances or can be used to explain phonological phenomena hitherto resistant to other explanations (W.H. Jacobsen, R. Austerlitz). Others examined those categories of words known as *ideophones* (G. Tucker Childs), *expressives* (G. Difloth) and *mimetic words* (S. Hamano). Most relevant to my own experiment are the pieces by Diffloth and Priestly because they explored the boundary between the language-specific and the universal in phonetic symbolism.

In examining expressives in Bahar, a Mon-Khmer language of Vietnam, Diffloth was concerned to dissociate simple ononmatopoeia from "the Bahar iconic system" (Diffloth 1994: 109). He came to the conclusion that "High = Big and Low = Small, exactly opposite to the English /i/ = Small and /a/ = Big, claimed to be universal.

"There is nothing peculiar about this Bahar system, and one can easily find an iconic basis for it. In the articulation[82] of high vowels, the tongue occupies a much larger volume in the mouth than it does for low vowels. The proprioceptive sensation due to this, reinforced by the amount of contact between the sides of the tongue and the upper molars, is available to all speakers and is probably necessary to achieve a precise articulatory gesture. The most direct form of iconicity relies on finding similarities between two

[82] Apparently, Diffloth does not consider the acoustic aspect of speech sounds as likely to trigger iconicity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

different kinds of sensations: articulatory feedback sensations... on the one hand, and the various sensations conveyed by expressives on the other.

"In this perspective, two languages may easily use the same phonetic variable (vowel height) to convey the same range of sensations (size) and come up with exactly opposite solutions, both being equally iconic; all they need to do is focus upon different parts of the rich sensation package provided by articulatory gesture, in our case the volume of the tongue instead of the size of the air passage between it and the palate. *Iconicity can be both physiologically motivated and culturally relative at the same time*" (Diffloth 1994: 113, emphasis added). This is exactly Werner and Kaplan's (1963) point about "plurisignificance of vehicular material" --that one cannot assume which *aspect* of the linguistic vehicle (in this case, a vowel) will be iconically matched to meaning.

Priestly made a very useful distinction between two types of phonetic symbolism: "synesthetic (or: universal) and conventional (or: local)," stating that they "co-exist in every language" (Priestly 1994: 237). However, in his equating *synesthetic* with *universal*, there is a conflict with Diffloth's view that in going from one culture to another, no sound-meaning correspondance should be taken for granted. The classic synesthesia, colored hearing does not manifest as the same relation of phone to color in every subject. (Reichard et al, 1949). Further, /i/ in Fischer-Jørgensen's (1987) account is generally seen as the brightest vowel by most English and French speakers—but not all, and for West African language speakers it is seen as black.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Quoting the work of Zuravlev (1974) and Levicki (1973) "both based on testing procedures that are difficult to assess" Priestly noted certain "apparent anomalies," for example, "that Russian /a/ = FAST for Russians, whereas Ukrainian /a/ = SLOW for speakers of closely related Ukrainian," remarking that "it is inattention to the probable coexistence of (at least) the two kinds of associations outlined above has bedevilled most of the discussions of sound symbolism" (Priestly 1994: 237—238). He suggested that two different analytical procedures should be used on the two different types of phonetic symbolism, since "synesthetic associations are necessarily 'more-or-less'" while "conventional associations are 'either-or.'" Thus "synesthetic associations should be analyzed on phonetic and semantic scales, while conventional associations should be analyzed in terms of distinctions between the phonemes or... the presence or absence of the distinctive features that compose them, and, similarly, in terms of 'plus-or-minus' semantic distinctive features." (Priestly 1994: 238).

He made a study of four poems by Voznesenskij in phonetic transcription, segmented into syllables (with calculation of syllable values through poem-specific syllable strength-weightings and additional strengths added to syllables due to previous analysis of the poems' sound structures, i.e. "foregrounding effects"). The study focused on the distribution of grave and acute segments (i.e. on a 'universal' phonosymbolic aspect), computing the amount of weighted syllables on a seven-point acute-to-grave scale and correlating the resulting values with judgements of the poems' emotional tones, specifically brightness/happiness versus darkness/sadness. Priestly's ranked analyses were later found to be "identical" with those of the responses of two small groups of listeners

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(Russian-speaking poetry specialists) who were asked to rate the poems on a ten-point sad-to-happy scale.

To conclude this section, I would like to remind the reader of Fischer-Jørgensen's (1978: 87) remark "Phonetic symbolism cannot be explained independently of the general phenomenon of synesthesia." This is clearly not true of secondary phonetic symbolism, and there are certainly some linguists (of whom Saka (1988) represents the most extreme case) who argue that language itself is the primary phonosymbolic generator. However, I believe that it is not possible to give a rounded account of this phenomenon without *some* reference to synesthesia experimentation and theory. Indeed, there have been a number of experiments in phonetic symbolism, matching words to drawn figures—the most famous being Köhler's (1947)—whose results are simply not understandable in terms of linguistic self-referentiality. Therefore, the next section will treat the theme of fittingness and synesthesia, attempting to locate phonetic symbolism in its larger context.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 4

PHONETIC SYMBOLISM: STUDIES OF FITTINGNESS AND SYNESTHESIA FROM 1934 TO THE PRESENT

Fox (1935) reported on a lengthy series of *gestalt*-focused experiments, but since all but one of his 10 tests was run with a very small number of subjects (5 or less), his results can have no more than anecdotal significance. However, his findings were to bring up some important points. Very aware of the work of his contemporaries, he prefaced his report with a substantial review of the literature, mentioning amongst others, Köhler whose influential *maluma/takete* word-figure quartet (Köhler 1929) he was perhaps to be the first to use in this country. Fox also cited a number of experiments available—still—only in German.[1]

Fox's investigation was "undertaken initially as a check upon Usnadze's" (Fox 1935: 548),[2] especially his findings on associational factors, although he was clearly somewhat

---

[1] He included ones by von Hornbostel "unfortunately not reported fully" (Fox, 1935: 546), by Fischer in 1922 and by Usnadze in 1924. Kohler cited Usnadze as an inspiration for the maluma/takete matches, but nowhere in Usnadze's (1924) report are these stimulus words mentioned.

[2] Usnadze had isolated four possible associational factors which might be responsible for naming a nonsense (i.e. non-representational) figure with an artificial word. In his experiment, he claimed that it was one of the three Gestalt-type factors which led to consistency in naming, rather than the fourth, which was lexical association. Thus agreement between the *form* of the word and the object, agreement evoked by the *feeling* of the word and the object, or agreement *linking word and object* in some less defineable way-- "a 'general impression' serving as a basis for the naming structure" (Fox 1935: 594)—were all factors that led

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sceptical about Usnadze's work. He criticized Usnadze's experiment as not being comprehensive enough in the amount of data collected, in having insufficient statisical analyses and inadequate description of stimuli. Moreover, he felt that Usnadze had accepted his subjects' introspections much too easily at their face value.

Fox presented his stimuli were visually,[3] using artificial words of the forms CVC, CVCVC or CVCVCVC and employing "all the letters of the English alphabet" (Fox 1935: 555). This sound-grapheme confusion unfortunately led to some phonic exclusions and duplications—i.e., to using by no means all the sounds of English.[4] Subjects were required to match the figures to the most fitting of the words on the cards and to provide introspective reports on their reasons for finding agreement between a word and a figure. Initially, there was a third task: to recall the names selected for a series of figures after a five-minute break, but this requirement was dropped when it was found to bias word-choice in favor of the most easily remembered items. Instead, two of the subsequent tests were repeated after two-month intervals with the same stimuli (in a different order) and the same subjects—who were not told that the tests were repeats.

---

to the same word's being chosen a second time as an apt name for a figure even though "the O typically did not know the word he had chosen before." (Fox 1935: 594).

[3] These were in the form of translucent cards which were set for ten seconds in an illuminated screen in front of the subject. Each card had on the left a simple non-representational figures "as distinctive as possible" (Fox 1935: 555), made up of four strokes, and on the right a short column of artficial words, listing 3 or 5 items. Because of the visual presentation (following Usnadze's procedures), Fox unfortunately introduced an extra variable: the problem of the visual influence of the letter shapes and the forms of the words that they composed.

[4] Thus the phonemes reprsented by the digraphs *th*, *sh* and *ch*, together with *eng* and *zh* were excluded. Pronunciations of *c* and *g* were standardized (as velars); on the other hand the redundant *qu* and *k* were also included. Vowel pronunciations were standardized to (only) five sounds.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In contradiction with Usnadze's findings, Fox discovered that throughout his whole series of tests, lexical associations were as productive of consistent recall as any other type of associative factor. He also observed early on that "nothing in the reports given by the other O's contradicts the *possibility* that many of the reported reasons are discovered or elaborated *after* the choice" (Fox 1935: 561).[5] However, he noted that for all participants there was one area of truly systematic attributions: "sharpness and angularity on the one hand and roundness, smoothness and voluminousness on the other" (Fox 1935: 570) Because of the visual stimulus presentation, this measure may have been particularly available to subjects, since angular/rounded applies well to the forms of letters. Based on his findings, he established a scale beginning with the letters most frequently characterized as "sharp" and ending with those least frequently so characterized.[6] Fox counted these findings as significant because they emerged without subjects' being given "any formal instruction to search for these particular symbolisms" (Fox 1935: 571), and because they were so clearly established despite the levelling effect of the use of the letters in many words that were not judged as displaying these attributes.

Following the repeats, a couple of further tests were run, firstly asking subjects to provide their own names for non-representational figures and secondly their own drawings for artificial words. Something of the problematic complexity of using whole words as stimuli was indicated in Fox's report on the second of these tests: "The most difficult

---

[5]This suspicion was borne out by later findings: "In this series (as in others) O's agree very little in the reasons which they give for choosing the same words... we find a correlation far from perfect between the word selected and 'reasons' for its selection" (Fox 1935: 570).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

aspect of the problem was for O to construct a figure which *as a total* was well named by the word *as a total*." (Fox 1935: 569, emphasis added). Fox found that there was "very imperfect agreement between O's as regards the 'symbolism' of particular sounds" and, apart from one subject, "even for one and the same O... not always a perfect consistency in characterizations reported" (Fox 1935: 570). After a number of further small, inconclusive tests, Fox ran his penultimate one with 78 subjects based on Köhler's *maluma/takete* quartet. On this, all but 4 subjects correctly matched figures and names. Apart from his *maluma/takete* findings, and those on attributions of roundness and angularity, Fox's most important discoveries were negative: that the non-lexical associative factors Usnadze claimed as motivating consistency in word-choice did not operate in case of his (Fox's) subjects, and that subject introspections were not a reliable guide to understanding how figures are named.[7]

More or less simultaneously, Irwin and Newland (1940) pulished a study that followed Fox's work matching non-representational figures to artificial words, but they used children as subjects. They reasoned, following William James, that similarity of perception is a sign of intelligence and so wanted to investigate this in "perception of configurational similarities in groups of normal children" (Irwin and Newland 1940: 4) Stimuli were in five sets, each comprising a pair of artificial words and a pair of non-

---

[6] The sharpest letter was *d*; *m* was the least sharp. *i*, judged as sharp 77.4% of the time, ranked well above *a* and *e* which clustered together at respectively 40.0% and 39.4%, which in turn ranked above *u* at 17%.

[7] "Our experiments were not adapted for a detailed study of the real bases underlying naming. The non-correlation between real and reported bases which we discovered so unexpectedly made a systematic study of real bases impossible... O's agree upon the same word in a particular case much more frequently than they agree upon their bases for selecting it. *It is therefore essential to distinguish between the real bases underlying the choice of names and O's reported reasons...*" (Fox 1935: 579, my italics).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

representational drawings, including Köhler's *maluma/takeetee* pair.[8] The experimenters themselves designed figures to correspond to the words, (but gave no criteria for or description of what constituted fittingness). Resulting stimuli were then piloted with 23 adult subjects with a 78% agreement on appropriateness of the names. However, several of the stimuli appear to have been less than perfectly designed.[9] Subjects were 306 schoolchildren, ranging in ages from 4 to 18. There were 146 boys and 160 girls, but no gender differences were discovered in the results which were therefore pooled.

The experimenters observed that they could not be sure (since a set contained four entities, two pairs of figures matched with two pairs of words) whether the presence of *both* names and figures influenced subjects' decisions, or if the subjects would have made the same choices with one figure or name absent.[10] But they claimed that "the sensitivity of the experiment is increased by having two figure-name pairs at the moment of choice" (Irwin and Newland 1940: 11). Presentation was done individually; the experimenter showed the subject two drawings and pronounced two names, asking the child which

---

[8] Every experimenter (beginning with Köhler himself at different times) seems to have spelled these words slightly differently.

[9] Set 3 included the word SKITZA; by age 13, *schizophrenic* might well have been in some subjects' vocabularies, and certainly by age 18 (the upper limit). The other word in set 3, LUN, is the central syllable of LALUNA of set 4, and LUN's accompanying spiral figure bears some resemblance to LALUNA's corresponding curled figure. Set 4, which drew the most positive responses,seems very easy, perhaps too easy, matching a single stroke drawing to the monosyllable LAH and a more complicated shape to the polysyllable LALUNA; in contrast, set 5 appears very confusing. Two monosyllables, JIJ and AK, are to be matched respectively to a cirular configuration of curving strokes like petals and a square with each angle marked with a diagonal line across it, giving the impression of a second square inside the first and at right angles to it. Any matching here would seem to be entirely arbitrary, and with their child subjects this set did indeed yield negative results.

[10] This issue was to reappear with word-matching tests. Brackbill and Little (1957: 314) referred briefly to a pilot study of theirs indicating that only one word (and not a pair) was necessay for matching with another

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

name corresponded to which drawing. Within each set the order of presentation of names and drawings was varied, but the five sets were always given in the same order. Irwin and Newland found that the very young children had no significant correct agreement apart from the fourth set.[11] They discovered a steep increase in ability to judge correctly between the ages of 5 and 9; responses to all the sets showed the same trends in correct judgement, a maximum of 88% possible after 9 with peaks at 9, 13 and 17.[12]

The subjects were also asked to give the reasons for their choices; 216 responses were obtained, which were classified as *attributive*, *general impression* and *associative* (i.e., lexically based). No especial pattern was noted between age and type of reason, except that no associative reasons were offrered before age 9 (also the age when correct matching rose significantly above a chance level on most sets). In contrast with Fox's findings, there was a significant majority (over 75%) of attributive, i.e., non-lexical, reasons for choices.[13] Irwin and Newland conceded that Fox may have been right and that the reasons given may have been no more than post-facto rationalizations, but still found it hard not to conclude that "a large number of their responses were based on perception of common properties in names and figures" (Irwin and Newland 1940: 14).[14] The most

---

on *same* vs *different* judgements, but no details of this pilot appeared in their article.

[11] The fourth set generally obtained the highest agreement with all ages, and 100% with the 18 year olds.

[12] This was with the exception of the confusing fifth set. The authors did not try to explain the lack of success with this set, or the opposite results with the fourth set.

[13] Those quoted in the general impression category greatly resemble the confident, matter-of-fact reactions Eberhardt observed with her first subjects aged between 6 and 8.

[14] They also suggested that since some subjects gave non-standard (i.e. incorrect) responses after giving standard ones, this showed that they understood the type of relationship that the experimenters were studying but that maybe "some of our material was not 'within the vocabulary' of some of our (child)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

significant aspect of this study is the strong positive subject reaction to the fourth set (matched correctly well above chance, even by the four-year olds). This set displays what has been suggested by Brackbill and Little (1957) and many subsequent researchers as the most salient and easy to grasp structural difference: word length matched to figure complexity.

The authors found little or no evidence to support their original hypothesis. Comparison of correct responses on the tests with I.Q. scores (used as measures of subjects' intelligence) showed no conclusive correlation, with the possible exception of the older children's responses, but they noted that this slight linkage could be due to to age, to the type of questions asked in the study, or to some other factor. [15] In contrast, evidence that the synesthetic faculty (and by implication, the faculty for phonetic symbolism) is *not* confined to a special "gifted" group was provided in two companion papers published a couple of years after Irwin and Newland's: Odbert, Karwoski and Eckerson (1942) and Karwoski, Osbert and Osgood (1942). Neither reported anything on phonetic symbolism per se, although the authors remarked that there were implications for the general study of psychology and of language.[16]

---

subjects" (Irwin and Newland 1940: 15).

[15]If Irwin and Newland had been supported experimentally in their hypothesis, this would have suggested that all forms of kinesthesia, including phonetic symbolism, are indicators of especial intelligence, rather than aspects of a normal and universal faculty. Capacity to focus conscious awareness on a sense of correspondances, to introspect and to articulate one's findings in detail may indeed be a mark of above-average intelligence (the literature on phonetic symbolism is strewn with good introspectors and articulators, starting with Plato). However, it is at least as likely that this sense, as von Hornbostel observes (1927: 84), is present in all of us: "Anyone can find on the piano the tone which sounds as bright as lilac smells. (Generally he thinks the task nonsense at first, but if he can be persuaded to deal with such nonsense at all, it goes very well)."

[16] See Karwoski et al. (1942: 204-205, 210, 213-214).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The first study investigated the connection between the mood of a piece of music and colors suggested by the piece; the second looked at the "form factor in colored hearing" (Karwoski, Odbert and Osgood 1942: 199). Despite variability of subject responses, the authors felt very strongly that their results showed that they were accessing a common synesthetic reaction. In reporting on the first experiment, they stated, "For some subjects the color experiences were quite direct and vivid; for others the judgements were forced. All types of judgement resulted in similar mood-color relationships, *suggesting the existence of some basically similar psychological mechanism in all people, regardless of the vividness of their imagery*" (Odbert, Karwoski and Eckerson 1942: 153, emphasis added). Subjects were 243 psychology students, who were asked to judge 10 musical phrases; first they rated each phrase on Hevner's (1936) 8-dimensional circular mood scale,[17] and and then reported if they *saw*, *thought*, or *felt* a color connection with the music.[18] Around 20% of subjects reported "consistently seeing colors" (Odbert, Karwoski and Eckerson 1942: 158). On the other hand, 33% of all reports were of judgements "forced" by the experimental requirements. The authors concluded, "subjects who are forced to relate color to music give responses very similar to those of subjects who react readily to music with visual imagery. The chief difference is in the richness of response" (Odbert, Karwoski and Eckerson 1942: 172).

---

[17] For a discussion of Hevner's work, see Chapter 5.

[18] There was also the possibility of opting out of a decision if no relationship was experienced between a phrase and a color. In experiments of this type, such an option is always a strength in the design.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In a companion study on the "patterning process of visual responses to music" (Karwoski, Odbert and Osgood 1942: 199) the performance of an unselected control group (100 college sophomores) was compared with that of a group of selected "photistic visualizers" (Karwoski, Odbert and Osgood 1942: 199), i.e., ones chosen for degree and vividness of colors reported on hearing music.[19] Findings were that there was "little qualitative difference between the drawings of the selected photistic visualizers and those of the unselected control group" (Karwoski, Odbert and Osgood 1942: 201). The researchers also found that a specific emotional response might be represented by an array of different drawings, all, however, clearly understandable as communicating the same thing. [20] A final test asked 100 unselected subjects to choose as a matching triad one member out of each of 3 pairs of polarized epithets representing three different areas of experience: visual, musical and emotional. Again, non-photistic subjects showed exactly the same tendencies as photistic subjects had earlier.[21] In a concluding summary, the authors stated that their results emphasized the "continuity between synesthetic

---

[19] This selected group drew their responses to musical selections; reactions were tested separately to musical structure (with 16 subjects) and to melody (with 21 subjects). The control group indicated responses on a printed form, but in the melody test, another control group of 12 was asked to draw their responses to the music"as best they could" (Karwoski, Odbert and Osgood 1942: 200).

[20]The drawings displayed "a variety of devices... to represent the same stimulus selections. However, the various 'choice' responses were all clearly recognizable as appropriate symbols for the stimulus selections. It would appear, then, that the various responses to a single selection are 'equivalent.'" (Karwoski, Odbert and Osgood 1942: 202).

[21] "The agreement (of these unselected subjects) with the relationships observed in photistic visualization is very high. The correspondance is so high that we may conclude that the *capacity* to appreciate music in visual terms exists in a considerable proportion of the population." (Karwoski, Odbert and Osgood 1942: 208). However, the experimenters discovered subjects' reporting a great range of different bases on which they related the items, concluding, like Fox, that reports should not be taken at face value: "A basic, common frame of reference may actually be operating, although it is not verbalized" (Karwoski, Odbert and Osgood 1942: 210).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phenomena and more general phenomena of language and thinking" (Karwoski, Odbert and Osgood 1942: 213).

Three interesting points arise from these studies; the first is the striking commonality of perception between gifted visualizers or colored hearers and normal subjects. Second comes the discovery of the great diversity of graphic forms, all "readable" as equivalent, which were used to communicate similar emotional perceptions in the visual part of the experiment. A third significant finding also emerges from this study: much of subjects' translation from the modality of sound to that of sight involved the use of *oppositions*: "translations occur most readily and commonly in terms of opposites such as high-low, fast-slow, thick-thin etc." (Karwoski, Odbert and Osgood 1942: 205).

The observations of Odbert, Karwoski et al. (1942), suggesting the relative normality of synesthesia, were given further support in a paper by Reichard, Jakobson and Werth in 1949.[22] The synesthetic introspections of a single informant, a speaker of six European languages, formed the central part of this paper, followed by commentaries.[23] During the authors' "short, informal investigations... it was found that some adults retain their early

---

[22] "We have found that these phenomena are not nearly so rare as is supposed... Several linguists who are in a position to record the development of children's learning aver that children take it for granted that sounds and words have color, but the child's assumption is either unrealized by adults, or if discovered may be laughed at" (Reichard, Jacobson and Werth 1949: 224).

[23] The authors observed that "some of the sounds... have form rather than color" and that "percepts are not due exclusively to sound, some combinations are determined by the letters..." (Reichard, Jacobson and Werth 1949: 227). They quoted Nabokov's description of his own case in which "the color sensation is produced by orally forming a letter while imagining its outline" (Reichard, Jacobson and Werth 1949: 227), i.e. of interaction between the orthographic and phonetic aspects of a letter-sound complex.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

childhood pictures and correlate them with other percepts" (Reichard, Jacobson and Werth 1949: 224), but that this was perceived by some informants as a childish and embarrassing regression.[24] When Jakobson discussed the matter with a psychiatrist, "remarking that it was normal to interpret sound in terms of color, the psychiatrist replied that this was only true of 'persons suffering from mental aberrations'... he could hardly believe it of normal individuals" (Reichard, Jacobson and Werth 1949: 224).

Nevertheless, the authors insisted that the materials which they presented showed "that some normal adults, probably many more than we realize, synthesize sounds and other percepts... *there can be no doubt that forms of synesthesia exist in many minds even though the details may differ*" (Reichard, Jacobson and Werth 1949: 224, emphasis added). Admitting that their materials "were obviously very uneven" and "correspondance is far from regular" (Reichard, Jacobson and Werth 1949: 230), the authors called for studies productive of more consistent and controlled data. Moreover, they conceded that other synesthetic aspects of sounds, "light, form, texture and motion" (Reichard, Jacobson and Werth 1949: 231), also required investigation. Although admittedly inconclusive and anecdotal, this study makes some important points about synesthesia. Especially striking is its suggestion that synesthesia is much more widespread in children than in adults[25] This in turn suggests that if synesthetic experience

[24] "They have always been ashamed of the specific 'pictures' and have therefore never talked about the subject" (Reichard, Jacobson and Werth 1949: 224).

[25] The implication is that it is a more or less unconscious faculty driven underground (and, presumeably, lost) through self-consciousness or shame: "It is probable that those children who are trained by sympathetic adults retain their color images into adulthood..." (Reichard, Jacobson and Werth 1949: 230).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is a childish substrate, given the right kind of experimental stimulus, it should still be accessible to observation.

In a speculative, introspective response to Reichard et al, Masson (1952) made an important distinction between "individual synesthesias, situational in origin, and general synesthesias, common to all or most of humanity" (Masson 1952: 39). Yet in developing this idea, his original dual grouping was interestingly reanalyzed as a *triad* of possible synesthesia types: the first individual, the second shared but depending on such things as accidents of geography, and the third truly universal.[26] Regrettably, it did not occur to the author that a speaker's native phonology, morphology and lexicon might have a systematic, non-individual yet not properly speaking universal effect on phonetic symbolisms.[27] Instead, while allowing for "individual and chance associations," Masson hypothesized an acoustically-triggered universal set of vowel attributions,[28] although he also credited non-acoustic influences with synesthetic effects.[29] Masson's speculative

---

[26] Such systems of correspondances as alphabetic and number synesthesias were seen as depending largely on individual experience, while synesthetic associations to time systems (producing such effects as consistently-colored hours of the day or seasons) would depend on accidents of shared geography. A third group, "the most interesting from a linguistic point of view, includes some of the true acoustic synesthesias, due to music or speech sounds" (Masson 1952: 39). The author suggested that despite "intrusions from private sources" into the associations of this third type, classic systematic colored hearing related to either acoustic or articulatory universals "to size and shape of the mouth cavity... or the pattern of formants and frequencies in both musical and speech sounds" (Masson 1952: 39).

[27] In fact, there is an argument to be made that one's native language is, like one's geographical location, a shared accident, and that some linguistically based synesthesias may belong to the second, rather than the third, group.

[28] He proposes a "natural and unconscious relation in the human mind between the distance separating formants 1 and 2 of a vowel and the luminescence of synesthetic coloring." Moreover, the central vowels, having "an even spacing over so wide an extent of the field of possible frequencies of formants 1, 2 and 3," this acoustic patterning "is directly connected with the muddiness of their synesthetic colors; tone saturation is symbolized by hue saturation." (Masson 1952: 40).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

conclusions all suggested a universal phonetic connection between colors and frequencies or colors and gestures, and apart from making no allowance for language-specific influences, appear quite plausible.

A study by Hall and Oldfield (1950) on the fittingness of signs to words was inspired by Bartlett's (1932) experiments, reported in his book *Remembering*.[30] The authors wished to examine "affective fittingness" (Hall and Oldfield, 1950: 61)—that feeling of appropriateness which, "*peculiar and unmistakeable... always occurs when some presented material seems to be singularly in its right place, but we do not know why it should be so*" (Bartlett, 1932: 113, emphasis added). In his study, Bartlett concluded that "fitness of certain material may involve complex cognitive processes which do not need to be formulated in order to be appreciated" (quoted Hall and Oldfield, 1950: 61). Observations of this type are quite normal when the making or experiencing of art is under discussion;[31] what is essential to note is that the feeling of fittingness is described as having a *pre-rational* quality. This note is struck many times in the literature on fittingness and synesthesia.

---

[29] "Emotional expression undoubtedly plays a part"... The "warm agressive tints" which for many people are associated with the low back and low central vowels "must be connected to some extent with a natural tendency in men and mammals generally to use these sounds in an expansive or fighting mood" (Masson 1952: 40).

[30] Bartlett found his subjects "divided naturally into two classes" with different styles of remembering (Bartlett 1932: 59). They were either "visualizers" or "vocalizers," i.e. those whose recall was triggered by language cues. This division suggests that there may also be more than one style of attributing phonetic symbolism, depending on one's personal predisposition either to focus more on aural or more on visual experience.

[31] The reader is reminded of Edith Sitwell's remark (quoted in Chapter 2) on her poetry as initiated without thought, by "automatic writing," and also T.S. Eliot's comment that true poetry communicates before it is understood.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hall and Oldfield's stimuli were six cards, each having one word and two "symbols or signs" (Hall and Oldfield, 1950: 61), i.e., drawn, non-representational figures. The experimental task was to assign the word to one or other of the figures as more appropriate, and to give reasons for the selection. The experimenters were interested in how far there would be consensus in the choice of a sign as fitting a word where subjects were from a similar social and intellectual background. They also wanted to look at the diversity of reasons given for a fitting choice, and, finally, to see if those signs chosen to fit words were better recalled than the signs that had not been selected. Forty subjects took part in the main experiment, and results showed upward of 75% consensus. Neither the position of the figures on the card (whether on the right or left) nor the sex of the subjects was shown to have any significant effect on figure choice. However, a preliminary experiment with slightly different procedures and less subjects (17) indicated that "the *general* attitude of the subject may excercise an influence on the *particular* choice made in respect of the character of fittingness" (Hall and Oldfield, 1950: 65).[32]

In examining the reasons for choices, the authors "gained the impression that the choice itself was a direct schematically determined reaction like the act of recognition, while the evocation of the reason was less a statement for the causes of the choice than a *justification* of an attitude already evoked in the choice" (Hall and Oldfield, 1950: 66). They concluded that consensus could be high, yet the reasons for a choice "very various

[32] This importance of attitude had been noted by Bartlett who distinguished between unfocused and focused perception (*hearing* versus *listening*, or *seeing* versus *noticing*) as influencing accurate recall: "determined mainly by qualitative differences of stimuli in relation to predispositions-- cognitive, affective and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in their wording... and even in the type of reason given" (Hall and Oldfield, 1950: 67), and "often... a rationalization of a more immediate perceptual process" (Hall and Oldfield, 1950: 70). They also noted that *contrastiveness* was an important factor in the sense of fittingness: "A sign is not absolutely fitting in connection with any verbal setting, but is selected in its comparative relation to the other member of the pair, making up the total problem-unit" (Hall and Oldfield, 1950: 67).

Between those figures originally chosen as appropriate and those later reproduced in a memory test, the experimenters found a high correlation—twice as many fitting as non-fitting figures were reproduced. The authors remarked that "verbalization[33] intervening as a reaction between the making of the choice and the making of the reproduction has in many cases influenced this reproduction" (Hall and Oldfield, 1950: 69), i.e., that the relation was not just between original word and figure but between word, figure and subsequent verbally-mediated attitude towards both in the reason for the selection. Hall and Oldfield concluded that in the chosen figure, a sense of diagrammatic iconicity (this is Peirce's useful term)[34] made for better recall.

Seven years later, Hall ran a follow-up experiment, where the experimental task involved a choice of five figures (rather than a pair), which were gradable in terms of their

---

motor"—of the perceiver (Bartlett 1932: 191—192).

[33] Also noted by Bartlett: "Naming (i.e.of figures reproduced from memory) strongly affects reproduction, whether it be immediate or remote" (Bartlett 1932: 185).

[34] In other words, the relations of the parts of the sign to each other represented similar relations in the iconized referent. (See Waugh, 1993, who describes Peirce's three icon types).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

abstractness or representationality. The experimental questions were whether the degree of abstraction of a figure would affect choice, and hence consensus, and whether there were individudal differences in selections and selection methods. A final question was on the "nature of the 'peculiar' fittingness Bartlett observed" and "its chief factors" (Hall, 1957: 21). Thirtyfour subjects were required to select the most fitting figure for a word such as *truth*, *life*, *atheism*, or *ambiguity*, i.e.,"abstract in its normal meaning" (Hall, 1957: 21). Fifty words of this type were tested.

For only 9 out of the 50 words were there agreements in figure choice of 50% or over: "No absolute and consistent grounds for popularity can be deduced from the type of symbol chosen" (Hall, 1957: 23). Individual subject consistency or variability in selection of figure type was also examined. 65% of subjects preferred signs depicting objects, while the remaining 35% preferred non-representational figures. Those selecting non-represenational figures appeared to base their choices on "an attitude... of rejecting the common and banal association" (Hall, 1957: 27). Hall discovered, as might be expected, that the more conventional, obvious and concrete the figure, the more often it was chosen; for example that 19 out of 34 subjects chose a drawing of a horseshoe as the most fitting symbol for "luck." However, he observed that in the third of his subjects who did *not* choose the obvious sign, a radically different temperament type or "mode of apprehension" (Hall, 1957: 30) might be at work. Quoting Piaget (1926), the author referred to that childhood egocentric stage in understanding words, "what he (Piaget) calls 'subjective synthesis' or 'syncretism'" (Hall, 1957: 30—31), and implied that those subjects choosing the less conventional signs might be acting due to a survival of this

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

early sensibility. "Many choose the more objective, public or conventional sign wherever feasible, but... others choose *because of the way a sign affects them personally* rather than considering whether or not such a sign has any meaning for other people" (Hall, 1957: 31). This suggests a connection with Sapir's (1927) distinction in language dimensions, the *referential* giving rise to the conventional and social use of language (and hence a choice of those signs already socially agreed upon, like the horseshoe) and the *expressive* originating much farther back in preverbal childhood (so leading to a more idiosyncratic, less conventional sign choice).

Hall hypothesized that two possible broad classes of personality type could be responsible for difference in figure choices. According to Wulf[35] a *comprehensive* personality reacts to and names the general impression of the whole of a simple design (which would lead to the choice of concrete, conventional figures), while an *isolative* type tends to analyze the design and focuses on "the objective accuracy of his perception," (leading to a less conventional, more abstract figure choice). Yet another way of classifying would be following Jung's[36] distinction between "introversive and extroversive attitudes" (Hall, 1957: 30); with the introversive types choosing the less conventional signs. Hall's report ended with the reflection that "the implication of Piaget's studies on children's language and thought have not been developed in relation to the thinking of adults... *further investigation of the pre-logical thinking of adults is clearly of importance*" (Hall, 1957:

---

[35] Wulf, F. 1922 Selection 10 in *A sourcebook of gestalt psychology*, W. D. Ellis (ed) (1938) New York: Harcourt.

[36] Jung, C.G. 1916 *Analytical psychology*, London: Baillière.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

31, emphasis added). It could be argued that all studies of fittingness and of phonetic symbolism are investigating this very area.

A subsequent study by McMurry (1959) introduced the Semantic Differential (Osgood, Suci and Tannenbaum, 1957) as a new tool to measure fittingness. As stimuli, he chose all of Hall and Oldfield's (1950) or Hall and Oldfield's (1957) words, and most of their figures. In the first of two tests, 20 figures and 10 words were used; 156 subjects were asked to select one out of a pair of figures as the most fitting for a word. Then, in a second test, 39 fresh subjects were given the same 10 words and 20 figures to be judged on 15 scales from the Semantic Differential. His reasoning was that "because visual signs may be rated directly on this Semantic Differential, the degree of similarity between the associative meanings of any given word-sign pair becomes measurable" (McMurry 1959: 311).

Results from the first test showed "statistically significant agreement of choice" (McMurry 1959: 312) as to what constituted a fitting sign for a word. On the second test, McMurry calculated the mean ratings for all the words and figures, and thus obtained 150 instances where ratings of the signs chosen as fitting could be compared to the mean ratings of words. In 112 cases, the mean rating of the sign considered appropriate was closer to that of the word than that of the non-fitting sign—a statistically significant result. He concluded "that 'fittingness' of sign to word is related to the degree of similarity of their connotative meanings as indicated by ratings on the Semantic Differential" (McMurry 1959: 312).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

McMurry's interest in applying the Semantic Differential to the problem of fittingness was typical of its era. It was during the period of the mid-Forties and Fifties that this set of semantic measures was being developed and used with enormous enthusiasm. An article by Osgood and Suci (1955) described its design, explaining that it originated in synesthesia research, with the experiment of Karwoski et al. (1942) quoted above: "It was found that the process of translating from musical stimulus to "visual" response, for example, *could be described as the parallell alignment of two or more dimensions of experience, each defined in terms of polar opposites... with translations occurring between equivalent portions of these related continua...* This process is not limited to rare synesthetic individuals, rather being quite general and consistent with standard systems of metaphor in the culture" (Osgood and Suci 1955: 325, emphasis added).

The authors described the "logical basis of the semantic differential" as locating all judgements, even extreme ones, on an "experiential continuum definable by a pair of polar terms" (Osgood and Suci 1955: 326) They also discovered, via techniques of factor analysis, that "many different experiential continua, or ways in which meanings can vary, are essentially equivalent and hence may be represented by a single dimension" (Osgood and Suci 1955: 326). In the words of Brown (1958b), "Factor analysis is essentially a technique for pulling out correlation clusters.... Scales that function as near synonyms produce a cluster of mutually intercorrelated jugements. The factor extracted from such a cluster represents the experimenter's effort to characterize the common quality of the intercorrelated scales" (Brown 1958b: 150—151). Thus, in examining attributions to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

stimuli, only a limited number of continua (scales) may need to be used since the same underlying semantic factors may be called into play on an apparently varied number of judgements.[37]

In experiments designed to analyze the factor-loading of individual continua or scales, the authors discovered several metadimensions, of which three, *evaluation*, *potency* and *activity*, accounted by far for the most intercorrelational clustering. However, they noted that the metadimensions should not be considered as one hundred percent semantically pure and separate; for example, scales representing the potency factor tended "to be contaminated with the evaluative factor" (Osgood and Suci 1955: 330). The large/small dimension was identified as highly loaded on the potency factory, while active/passive was "distinctly loaded"(Osgood and Suci 1955: 332) on the activity factor, as was bright/dark.[38]

In *The measurement of meaning*, large/small is again identified as "dominantly representative of the potency factor" (Osgood, Suci and Tannenbaum 1957: 62). Dark/light[39] is also described as "including considerable loading on this factor" (Osgood, Suci and Tannenbaum 1957: 63), while active/passive is identified as a "pivot" or scale

---

[37] However, in an article reviewing Osgood, Suci and Tannenbaum's *Measurement of meaning* (1957), Brown (1958a) cautioned that factor-loadings are not absolute, but can change with context. In other words, they are sensitive to the object being evaluated: "e.g. whereas deep-shallow is not heavily loaded with evaluation for a miscellany of concepts, it is so loaded when sonar signals are scaled" (Brown 1958a: 114).

[38] They added that "the noticeable tendency for both activity and power to be associated with positive evaluation (e.g. *good, strong, active* go together rather than *good, weak, passive*) is probably a cultural semantic bias" (Osgood and Suci 1955: 334).

[39] The authors did not include dark/bright.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

most representative of the factor "oriented activity" (Osgood, Suci and Tannenbaum 1957: 63). Again, the authors noted some contamination: "scales representing this (activity) factor are not as independent of other factors as would be desirable, all tending to be somewhat positive in evaluation and potency." (Osgood, Suci and Tannenbaum 1957: 63).

Brown (1958a: 114) made the point in his review that "differential research is much closer to the literature on synesthesia and metaphor than it is to the semantic analysis of logicians." A whole chapter of his contemporaneous book *Words and things* (Brown 1958b), was dedicated to phonetic symbolism and metaphor, and included more than a passing reference to synesthesia. Surveying phonetic symbolism experiments to date, his conclusions were cautious and conservative.[40] In discussing metaphor, more specifically the metaphoric extension of sense experience, as instanced in expressions like *a cold face* or *a cold manner*, he suggested two possible explanations, the inter-sensorial and the mediated associational. His explanation of the first is as follows: "The principal sensory dimensions of the world are the same for men everywhere and are named in all languages. Though each of these dimensions is associated with one receptor system, its essential quality is inter-sensory. The quality is first detected in one sense-modality and is named at that stage. Afterward the quality is detected in many other phenomena that register with other senses. *The original name tends to be extended to any experience manifesting the criterial quality.* And so it happens that unrelated languages extend their vocabularies of

---

[40] "No sound-meaning correlation has been satisfactorily demonstrated to exist in the lexicon of any language... On the other hand, it is probable that most speakers of a given language find similar symbolic

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sensation in similar fashion." (Brown 1958b: 148-149, emphasis added). This account would agree with Grammont's (1933) explanation of phonetic symbolism as originally aural (onomatopoeic *imitation)* and then later extended (expressive *translation*). It also looks forward to Werner and Kaplan's description of physiognomization (see below) as essentially multi-sensorial.

However, there is a another possibility. Brown's second explanation distinguishes between *direct* (as in the general description given immediately above) and *mediated* associations, and suggests that the latter may be responsible for metaphor. "Many substances as they grow colder (thermally) also grow rigid and stiff and slow-moving, as witness molasses, honey, putty and the like. Suppose now that the movements of an unfriendly person also manifest stiffness and resistance to movement... Such people might be called cold because their behavior is stiff and stiff substances are often cold. This is a mediated rather than a direct association. Thermal cold is assumed to be a dimension of experience linked with the cold receptors in the skin. However, the activation of these receptors is correlated (though imperfectly) with other kinds of experience. This correlation is the ground for extending the world *cold* to new experiences that do not activate the cold receptors. The notion that sense qualities are correlated in Nature is central to the hypothesis of mediated association. There is some evidence of such correlation in factor-analytic studies of semantics" (Brown 1958b: 150).

---

implications in an unfamiliar phonetic sequence" (Brown 1958b 129-130).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Of course, these two accounts of metaphor formation are not mutually incompatible with each other. Clearly, much experience of the world (what Brown calls "Nature") would reinforce intersensory connections; for example the observed slow-movingness of chilled, viscous substances could connect with personal experience of inability to move swiftly and of bodily stiffness when the experiencer feels extremely cold. However, it can be argued that the capacity to notice and generalize about natural connections (such as the connection between cold and rigidity, or the fact that small objects tend to make higher-pitched noises than large ones) might well be preceded by a more primary stage of physical and emotional experience as a *gestalt*, in which the senses only gradually differentiate themselves—and in which the tendency to re-cluster, i.e., to return to the primary mode, is never far away.

In another cautiously enthusiastic review of *The measurement of meaning*, Carroll (1959) observed that several more metadimensions might be needed than the three major ones noted (Carroll 1959: 69—70). He also objected to the fact that all scales were adjectival and all concepts nominal, i.e., the research technique made use only of a modifier-modified construction. However, he considered that the authors' evidence for the metadimensions "would be sustained in further studies" (Carroll 1959: 73). Taking a behavioristic line, he suggested that they might originate in direct response to "the adjustment of the individual to the objects in his environment" (Carroll 1959: 73). Thus, "evaluation" corresponds to approach/avoidance of the stimulus, "activity" to the rate of adjustment to the stimulus and "potency" to the amount of adjustment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Carroll also observed that the Semantic Differential, "far from making the measurement of meaning more precise... capitalizes on ambiguity and vagueness. Both concepts and scales are like the many-armed deities of India: when they are put together, they can interlock in a multiplicity of ways. Notice, for example, how the adjective *hard* changes its meaning when linked with such concepts as ABORTION, ANGER, BOULDER, CRIMINAL--to sample a few which... have high ratings on *hard-soft*... Perhaps a certain degree of ambiguity in the experimental situation is requisite for teasing out semantic dimensions possessing a sufficient degree of generality. Indeed, all the well-known figures of speech—simile, metaphor, personnification—feed upon ambiguity, and one may suggest that Semantic Differential dimensions may be useful in describing and accounting for the uses of metaphor... The Semantic Differential was a natural extension of Osgood's prior studies in synesthesia" (Carroll 1959: 75—76).

The Semantic Differential was put to just this use in a series of investigations by Solomon (Solomon 1958, Solomon 1959a and 1959b) on a problem of pure (nonphonetic) sound symbolism. Solomon was examining attributions of qualities to sounds: passive sonar signals. His first subjects were 50 experienced submarine sonarmen, who already, like musicians, had developed a "loose and variable" professional vocabulary to describe highly significant sounds (in this case, those of different vessels) which they encountered. Solomon claimed the sonarmen were operating on the "basis of a semantic frame of reference" which was meaningful but "incompletely verbalized." He set out to try to "structure this frame of reference explicitly" by "providing... the development of several

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

standardized rulers and these rulers would be the dimensions of the frame of reference of the experienced sonarman" (Solomon 1958: 422).

Fifty scales from the Semantic Differential were used; stimuli were 20 sounds, 5 from each of "the 4 major catgories as differentiated by the Navy... warships, light-craft, cargo ships and submarines" (Solomon 1958: 423). Factor analysis indicated that eight different factors (i.e., possible metadimensions) were in play, of which Solomon specified the first seven as *magnitude*, *esthetic-evaluative*, *clarity*, *security*, *relaxation* and *familiarity*. He also noted that 15 of the scales used appeared to be "functionally pure," i.e., unambiguously loaded with one factor, and claimed that his technique "could be applied to any set of auditory stimuli, judged by any sample of the population" (Solomon 1958: 424).

In a second experiment, "an attempt was made to correlate physical measurements of the sounds with their rank orders on the seven psychological dimensions isolated by factor analysis" (Solomon 1959a: 492). Using data from his previous study, Solomon made acoustic analyses of the 20 stimuli, focusing on "the average sound-pressure level in each of the (8) octave bands from 37.5 to 9600 cps." (Solomon 1959a: 492-493), finding that certain dimensions, notably magnitude, clarity, security, familiarity and mood correlated fairly directly with energy in a given octave band or bands.[41]

[41] The first dimension to be analyzed, involving judgements of "heavy or high magnitude" (Solomon 1959a: 494) showed a significant correlation with energy concentration in those 3 bands spanning 150 to 1200 cps, "another way of saying that when sounds have a preponderance of energy in the lower frequencies, they are said to be 'heavy' 'large' 'rumbling' 'wide' and 'low' (Solomon 1959a: 495).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From results on those 15 scales discovered to be "factorially 'pure'" (Solomon 1959a: 496) or "pivots" (Osgood, Suci and Tannenbaum 1957), Solomon devised another correlational measure, which showed that 19 out of the 20 sounds clustered into two large groups. The determining factor appeared to be beat pattern, with one cluster characterized by light beats and the other (divided into three subgroups) by either regular or irregular heavy beats. A third study, using 20 naive subjects with no sonar sound experience, attempted to "clarify the role of beat rate and accent" (Solomon 1959b: 986). Stimuli were 21 especially constructed "abstract sounds" (Solomon 1959b: 990), systematically varied along the three dimensions of spectrum, beat rate and accent. Although the latter two features did influence subject judgements, interacting with each other and with the first, once again the major attributional clustering of sounds was determined "on the basis of band pass condition: high frequency vs. low frequency" (Solomon 1959b: 986). In this, Solomon's findings agree with much of phonetic symbolism research, from Sapir (1929) onwards.

The esthetic-evaluative dimension did not present any systematic relationship with acoustic energy distribution; Solomon explained this as due to the idiosyncratic nature of this type of judgement reflecting "the unique experiential background of the perceiver" (Solomon 1959a: 495). (It was precisely considerations of this type which ruled against using a semantic scale with high evaluation factor-loading in my own experiment). The factor of clarity appeared to show "a trend from negative correlations in the lower octave bands to positive correlations in the higher octave bands" (Solomon 1959a: 495), while security showed trends "remarkably similar to the curve of sensitivity for the normal ear" with "psychological 'intensity'" increasing with acoustic energy until 1200-2400 cps and thereafter tending towards "'mild' and 'gentle'" for 2400-9600 cps (Solomon 1959a: 495).

With the relaxation factor, "sounds with energy-concentration in the octave band 600-1200 cps are definitely judged as 'tight'" (Solomon 1959a: 495). Familiarity judgements indicated that sounds with predominatingly low-frequency energy concentration (75-600 cps) were "strange"; "familiar" sounds had energy in the higher octave bands. (Solomon 1959a: 496). The final factor to be analyzed, mood, showed a correlation between "colorfulness" and low frequency energy concentration (75-150cps) and "colorlessness" and high frequency concentration (1200-4800 cps) (Solomon 1959a: 496).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Semantic Differential was also used by Wertheimer (1958) in a modest study of phonosymbolic fittingness in words. His aim was "to obtain quantititative measures" of the relationship between especially fitting-sounding words and their meaning, and "to find some functional correlates which... increase our understanding of how such fittingness operates" (Wertheimer 1958: 412). He devised two fittingness measures based on the hypotheses that more fitting-sounding words would score higher on the Semantic Differential scale than those judged as less fitting, and that "fitting words would take longer to suffer loss of meaning than those judged as less fitting" (Wertheimer 1958: 412).

Stimuli were two five-word lists of fitting and non-fitting words.[42] Arguing that a fitting word "must have certain 'tertiary' qualities... clear-cut characteristics" (Wertheimer 1958: 414), 10 scales were chosen from the Semantic Differential, including angular/rounded, tense/relaxed and active/passive, to measure such qualities. Subjects were 15 college students.[43] Presentation of words was carefully controlled for position effect (known to be strong) by alternating fitting and non-fitting words to be judged on the semantic scales. Results showed that non-fitting words registered much closer to the mid-point of scales than fitting. The relation of lapse of meaning to fittingness was also tested: words were rated for what point under repetition meaninglessness would set in.[44] Wertheimer found

---

[42] These had been culled from two original lists of 100 well-fitting and 100 badly-fitting words, filtered successively through 3 sets of judges.

[43] They had been selected from a larger group of 24 for having perfect agreement with the judges' ratings of good or poor lexical fittingness

[44] The assumption was that meaning lapse would take longer for a word whose *trace complex* (made up of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the median time of meaning lapse was one third longer for fitting than for nonfitting words, but there were great individual differences in lapse time for the non-fitting words. In concluding, the author admitted that since the number of words tested and the number of subjects were so restricted, that he could not provide definitive answers to his experimental questions. Moreover, he pointed out that there were syntactic differences between the fitting and non-fitting words used in his study; thus, "to check the present findings, additional words should be used," (Wertheimer 1958: 415).

The effect of context on Semantic Differential judgements (more specifically that of the syntactic category of stimuli to different scales) were observed by Weiss (1963b) in a follow-up to Wertheimer's experiment. Weiss was looking to assess the possible contribution of phonetic symbolism to perceived intensity of word meanings[45] where the experimenter had taken into account the factor of the emotionality or expressiveness of stimulus words. Thus, instead of Wertheimer's single measure, Weiss used two preliminary groups of subjects to rate his word stimuli on *two* measures, fittingness and emotion. In this way, he selected 43 words from an original list of 127, obtaining four groups: high fit and high emotion (as in *howl*), low fit and high emotion (as in *bride*), low fit and low emotion (as in *barn*) and high fit and low emotion (as in *glide*). A third group of 45 subjects then rated the words on 10 Semantic Differential scales, chosen to

---

sound, meaning and the connection of both) were more cohesive, and that cohesiveness was enhanced by fittingness.

[45] Weiss saw two possible explanations for marked lexical expressiveness. One was the form-first argument: that there had been an (unexplained, presumably originally unmotivated) occurrence of phonetic symbolism in a word which eventually "would reinforce meaning characteristics." The other asserted that meaning came first, claiming that "highly charged expressions" would "include greater gestural participation" so that such words would be "richer in phonetic symbolism" (Weiss 1963b: 624).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

represent the metadimensions of Evaluation, Potency, Activity, Novelty and Tautness. Effects of order of presentation were controlled for. Results presented "two major findings" (Weiss 1963b: 626). Firstly, in examining low-emotion versus high-emotion words, two out of the five metadimensions provided unexpected results. Tautness scores were in the reversed direction (i.e., with low emotion words obtaining more extreme ratings), while Activity results yielded identical mean scores for low- and high-emotion words. Secondly, with well-fitting versus badly-fittng word scores on Evaluation and Novelty, it was the *badly*-fitting words which elicited "*significantly more extreme* ratings" (Weiss 1963b: 626), although Potency and Activity yielded scores in the expected directions.

Weiss argued that the reason for these surprising results lay in the syntactic makeup of the stimulus words; he claimed that the Semantic Differential scales were sensitive to the fact that 5 of the 10 badly-fitting words were nouns, while all 10 of the well-fitting ones were verbs. Thus the metadimensions of Potency and Activity, "sensitive to the action qualities of the concepts" (Weiss 1963b: 627), would respond well to verbs, while those of Evaluation and Novelty would be "equally applicable to nouns and verbs, and thus show mainly the effect of emotionality, and... the (above-mentioned) puzzling reversal or 'interference effect'" (Weiss 1963b: 627). However, two thirds of Weiss' syntactic judgements on these stimuli are questionable.[46]

---

[46] In fact, 3 out of the 5 badly-fitting 'nouns' and 7 out of the 10 well-fitting 'verbs' were syntactically ambiguous (*root, route, rust, howl, whip, groan, roar, tap, glide, scratch*), so some kind of additional syntactic evaluation test would have been needed to ascertain how subjects classified them. The interaction between stimulus and scales certainly could have used more clarification here.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Even so, his conclusions are still interesting, because they emphasize the role of context. He asserted that extremeness of ratings was due to "connotations or emotionality" of the stimulus words, and that "the meanings of the words themselves in the present study constitute the 'context' in which the judgement of whether a sound is 'appropriate' can be made... One must know to what the sound is supposed to be appropriate before a meaningful judgement can be made... *Phonetic symbolism cannot be seen in any sense as having an existence prior to, or independent of, meaning categories, but must instead be considered a psychological response initiated or suggested by (word) meanings, which then allows such meanings to become attached to otherwise meaningless sounds...*

"Evidence for this... may be drawn from observation of such language phenomena as poetry *in which the evanescent appropriateness of sound is highly dependent on the total context in which it occurs...* Phonetic symbolism, then, is highly dependent upon suggestion, a fact which was very early noted by Bentley and Varon" (Weiss 1963b: 628—629, emphasis added). Therefore, Weiss further argued, "words are not inherently high or low in phonetic symbolism until the suggestion is made, or a context implies, that such a sound-sense relationship is present"(Weiss 1963b: 629).

A study of word-figure fittingness was undertaken by Davis (1961) who used Köhler's original (1929) *baluma/takete* (by then, *uloomu/takete*)[47] quartet of words and drawings

---

[47] *Baluma* was later altered to *malumba* (to avoid association with 'balloon') and then to *maluma*, Kohler's final (1947) form. In Davis' experiment it was once again readapted as *uloomu* in order to avoid confusion

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

as stimuli with a group of Kitongwe-speaking[48] Tanganyikan children aged 8 to14, entirely uncontaminated with contact with English. As a control, the same investigation was also done with two groups of English children, 163 East London secondary school students (aged 11 to14) and 118 secondary-schoolers from Oxfordshire (ages 13 to14). In order to accustom the children to the experimental situation, an introductory task was given of assigning sex to two figures of 'pinmen,' one square-headed and the other round-headed. Then the children were asked to label the figures appropriately "man" and "woman." The main experimental task was to copy the Köhler figures and assign names to them. In one condition, the names *takete/uloomu* were presented visually[49] and aurally; in a second condition, the names were presented aurally alone. The English control groups followed the same procedures, except name presentation was aural only.

Results on the introductory experimental task were generally in the expected direction with all three groups: the square-headed figure was "male" and the round-headed, "female." However, in the case of the East London boys, significant results in the opposite direction showed the way in which context may affect subject judgements. This group of boys assigned names unexpectedly, because "square" in the Americanized London slang of the time meant "unacceptable outsider" and in the view of the pubescent

---

with *malume*: 'mother's brother' in certain local dialects.

[48] Experimental instructions were actually given in Swahili, the lingua franca of the region, which the children were learning at school. The fact of being experimentally directed in a second language may have meant that some of the girls who were "much more withdrawn than the boys" (Davis 1961: 265) did not understand the intstructions.

[49] Fox (1935: 570) gave results on roundness/angularity which suggested that visual presentation of words had notable influence on evaluations.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

male subjects, this was assimilated to "female." On the main task, results for the English control groups showed "overwhelming evidence of the uniformity of 'takete' to the angular figure in the control groups," while "the African children match[ed] names and drawings in the same way as the English controls, thus providing strong evidence that it is the relation of appropriateness between the actual speech sounds and the perceived shapes which determines the choice, rather than the meaningful associations with specific obejcts" (Davis 1961: 264—265).

By using both vowels and consonants in the two words that reinforced contrasting impressions of angularity on one hand and roundness on the other, Köhler had designed stimuli that were particularly appropriate for a matching experiment. Davis' study, modest though it was, provided clear support for the view that Köhler's figure-word matching task was indeed cross-culturally understandable—if not universally so. This therefore seems an appropriate point to begin an account of a work that truly attempts to explain the nature of what Köhler and Davis were trying to pin down.: Werner and Kaplan's *Symbol Formation* (1963).

Although it is almost never referred to in the literature,[50] this work offers an intuitively plausible, coherent, *simple* account of the function of synesthesia in general and phonetic symbolism in particular in the genesis of language and human communication. It attempts to give a reason (other than that of being a mere primitive linguistic or pre-linguistic

[50] It is mentioned very passingly in Ervin-Tripp and Slobin (1966) and cited in the bibliography of Marks (1978); these are the only references that I have found to date.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

survival) for the phonosymbolic faculty, which is presented as an essential part of normal language development and use. Werner and Kaplan's book should be seen, along with the works of Darwin (1872) and Grammont (1933), as giving more insight into the true nature and origins of phonetic symbolism than almost anything else to be read on the subject. Its authors also provide an all-too-convincing explanation of why most phonetic symbolism investigations have misconceived the issuses, asked the wrong experimental questions and in consequence searched for unfindable linguistic phenomena. Perhaps this is why their ideas have been so consistently ignored in the literature.

According to the authors, there are two possible explanations of how a child learns to use words. One approach is that of the "label theorists" (Werner and Kaplan, 1963: 110) who feel that a child learns through repetition of adults words and speech patterns. The authors reject this explanation in favor of one which sees the very young child as responding in a global, undifferentiated, *organismic* (a favorite word of theirs) way to inner and outer stimuli, producing noises and gestures which gradually develop communicativeness. Thus, the child begins in a world of "things-of-actions articulated in terms of affective-sensory motor patterns..." As a "directedness towards knowing begins to emerge," the child's world changes from things-of-action to objects-of-contemplation... These [expressive] patterns thus undergo a *shift of function*... become the means by which he comes to know objects, to reflect on them, to present them to himself..." Related to this "shift from outward reaction towards inward reflection is the *internalization* of sensory-motor patterns... objects are given form, structure, and meaning through inner-dynamic schematizing activity which shapes and intertwines the sensory, postural, affective and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

imaginal components of the organismic state" (Werner and Kaplan, 1963: 18). This functional shift in the use of expressive patterns—from reactive to referential—recalls Darwin's (1955) first principle of gestural symbolism, of serviceable associated habits, whereby e.g., head-scratching, having literally meant irritation in the scalp, comes to mean mental perplexity.

The authors stress that this early internalization activity—in which the form-meaning relation is organismically based, and in which an iconic, literally representational element is *always* present—underlies all use of language. They describe the formation of referential objects as starting from "a primordial matrix... of affective, interoceptive, postural, imaginal elements... channelized into full perceptual articulation by the schematizing activity" (Werner and Kaplan, 1963: 18). The more familiar the object, the less time needed for the process of (perceptual) articulation, but however automatic object-perception seems, "the articulating (form-building) activity leading to a meaningful percept never drops out" but "remains linked to an underlying process of schematization. *If this bond breaks, meaning is lost*" (Werner and Kaplan, 1963: 18—19, emphasis added). And later: "Were such dynamic processes to be eliminated or impaired—such processes involving postural-affective-dynamic schematization—*the conventional forms would undergo 'loss of meaning*'" (Werner and Kaplan, 1963: 208, emphasis added).

In terms of the discussion of phonetic symbolism, the most interesting is their account of the progression from the very young child's "natural symbols" (Werner and Kaplan,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1963: 89) to full adult speech. The infant's pre-communicative excitement begins gesturally and phonetically as activities which will be gradually refined and specialized into pointing movements and the utterance of syllables; however, to begin with, the infant "strains towards objects and emits 'call-sounds' which are part and parcel of the bodily directedness..." At this stage, there is a "global unity of sound and bodily movement." (Werner and Kaplan, 1963: 77). These activities are the "precursors of depictions, and eventually the depictions emerge from such dynamic sensory-motor formations" (Werner and Kaplan, 1963: 87). One non-vocal example quoted was the child's imitating the opening and closing of the experimenter's eyes by opening and closing hands, mouth or eyes; another was mimicking a flickering electric bulb by fluttering eyelids: "The imitative expressions have developed into truly *pictorial* or *iconic* representations... the child, by means of sensory-motor patterns creates what have been termed 'natural symbols'" (Werner and Kaplan, 1963: 89).

Four experiential requirements must be met for representation to take place, so that eventually "non-imitating sound forms," i.e., the non-iconic, arbitrary words of adult speech, can be "endowed with significance" (Werner and Kaplan, 1963: 19). The first is the actual or potential "expressiveness of objects... the experience of expressive features in things seen and heard." Crucially, this "non-representational construing of objects as expressive is basic, and genetically prior, to the use of expressive properties in representation"[51] (Werner and Kaplan, 1963: 19—20). The second is the "transcendence

[51] For a fascinating discussion on the origins of language, in which the endowing of objects "neither dangerous nor useful in reality" with expressive meaning is seen as fundamental to its development, see Langer (1942: 110—118).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of expressive qualities" (Werner and Kaplan, 1963: 20), the awareness that the same expressive features can characterize otherwise dissimilar objects. In this awareness lie the roots of metaphor. The third requirement is the "intentional act of denotative reference," that is the child's conscious focus on using "one item of experience to denote another" (Werner and Kaplan, 1963: 20), the communicative impulse. Fourth and finally comes the "establishment of semantic correspondance," which is the "construing of the pattern qua *vehicle*[52] so that its expressive meanings correspond to the connotational structure of the referent" (Werner and Kaplan, 1963: 20, emphasis added).

Evidently, these conditions have already been met (in non-phonetic form) in instances such as when the child flutters her eyelids in imitating the flickering bulb. Yet the "natural symbol" marks only a very preliminary stage, in which the child's utterance, and what it refers to, have more or less experiential unity. However, out of the child's use of natural symbols there comes a communicative progression which is characterized by increased "distancing" (Werner and Kaplan, 1963: 99) between vehicle and referent. It is divisible into three stages of linguistic representation: "(1) naturalistic-onomatopoeic depiction, (2) physiognomic depiction, and (3) conventional representation" (Werner and Kaplan, 1963: 100). In their depiction of the first two stages, the authors echo Grammont's observations on the core-penumbral nature of phonetic symbolism in which a central body of onomatopoeic "imitations" necessarily precedes those metaphorical

---

[52] This is the authors' general expression for a (verbal) signifier; at different stages in the child's linguistic development, this may be termed a *vocable*, *monoreme* or *word*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

extensions or "translations" which constitute true phonetic symbolism.[53] Likewise, in Werner and Kaplan's account, the early stage of onomatopoeic imitations gives rise to a middle stage of physiognomic depiction, using translations derived from the preceding onomatopoeic stage.

Physiognomic depictions are "articulated vocal patterns which depict *other* than the sonic properties of events... a *visual* property such as size may be depicted linguistically by translating variation of length into variation of sound duration... the vowels *i* versus *u* may indicate 'high' versus 'low' or 'far' versus 'close,' etc..." (Werner and Kaplan, 1963: 104) Since this second-stage communicative activity involves more distance between the expressive vehicle and the referent to be expressed than with first-stage onomatopoeic utterances, it indicates an advancement. The authors base their opinion that language goes through a highly synesthetic period for all children on the view that "translation of non-sonic properties into sound-patterns (is) based on similarities derived from the primordial unity of the senses; qualities such as 'long' or 'short' are organismic rather than merely sensory phenomena" (Werner and Kaplan, 1963: 104—105).

A more advanced stage of distancing between referent and vehicle is entered when the child begins to use the conventional forms heard around him or her. But between the second and third stages, there may be a number of transitional word-forms used. For example, physiognomic or even onomatopoeic expressions may be retained but assimilated into the conventional language code, adding verb suffixes and so forth to

[53] For discussion of this, see Chapter 2 and Grammont (1946: 403-412).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

them. Or names which are composite, half "infantile-idiomatic, the other conventional" (Werner and Kaplan, 1963: 108), or combinatory, such as 'moo-cow' may persist until the infantile part is finally dropped. Or a more bilingual pattern is followed in which the child uses the earlier form in one social situation (with, for example, the mother) and the later form in other, less intimate, situations.[54] However the transfer is done, one would speculate that the most intimate and emotionally charged vocabulary would be the last to be made over into adult speech and the first to resurface as an infantile form under stress: words for urination, defacation, intense feelings and smells, body parts, certain important foods, endearments, diminutives, curse words and so on. Consequently, where markedly "babyish" words have survived into adult speech, we should look first there for obviously physiognomizable, or physiognomized, forms.

Thus, it must be re-emphasized that the authors see this early, necessary physiognomization of language as *continuing into and underlying mature language use*. One adult survival of this substrate quoted is "word-realism," or the magical-religious fusion of word and referent in the use of spells and invocations, based on the fact that "material comprising the vehicle is structurized to possess or embody those features which are apprehended in the object." This allows the practitioner to "treat both vehicle and referent as identical objects or objects belonging to the same class..." and, using sympathetic magic, to affect one through invoking the other. Moreover, "the whole

---

[54] For work touching on the issue of social domains and bilingualism, see Rosenblum and Pinker (1983).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

postural-motor activity involved in articulating the linguistic forms endows them with the potency of real objects or actions" (Werner and Kaplan, 1963: 36-37).

The authors also caution us against thinking we are too civilized for word-realism by citing the case of invectives and words of love as having "substantialized character" (Werner and Kaplan, 1963: 37). Another instance of substrate survival which they mention is the original use of euphemisms, and the need for replacement euphemisms—both are indications of how words become fused with tabooed referents. To their examples could be added three more mixed ones (i.e., also clearly influenced by lexical associations). Firstly, there is the writing and reciting of poetry, especially lyric, where brevity and concentratedness focuses poet and reader very much on phonetic qualities.[55] Second is the naming of children, and third (shamefully but predictably in our society) the naming of commercial products.[56]

Werner and Kaplan also argue—for our purposes, crucially—that under the right experimental conditions, *this physiognomizing substrate can be activated.* Describing an early experiment by Werner, in which adults were asked to physoignomize conventional words, they report no difficulty in getting subjects to "easily envisage the meanings in tangible features of the vehicle... the ease with which the subjects could physiognomize the linguistic forms suggests that such an apprehension of words is not a rare or curious

---

[55] However, the syntax of a poem may also play a mimetic (i.e. iconic, physiognomizing) role.

[56] Advertizers are extremely interested in and aware of the uses of phonetic symbolism in the naming and selling of products. (See Sound symbolism and naming (1996) a white paper published by Master-Mcneil, a company "totally dedicated to the resolution of naming and branding issues.")

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phenomenon, permanently discarded by adults; *rather, it indicates that physiognomic tendencies operate covertly in normal adults and can be rather quickly realized in perceptual material through experimental instruction.* Not one of the large number [30] of subjects... was unable to physiognomize the linguistic forms, and with but one or two exceptions, all the subjects found the task a quite natural one" (Werner and Kaplan, 1963: 38, emphasis added). Or again, they observe that beneath conventional adult language is "organismic activity which involves the putting into play of primordial operations... often overlooked... because... they are subordinated to, and integrated within, higher levels of functioning. *Nevertheless, they can be brought to the fore... under special experimental conditions*" (Werner and Kaplan, 1963: 206—207, emphasis added).

Moreover, the authors emphasize that physiognomic apprehension of words is multi-sensory, "carried not only in the modulated properties of sounds, but in visual, auditory, tactual and kinesthetic qualities... Meaningful vocables are compounds of total organismic activities (inner sensory-motor gestures) and the physiognomic aspects need not be, and typically are not carried solely by the visual or auditory components of such total organismic activity." (Werner and Kaplan, 1963: 223). They state that suffixes, such as the German agentive {-er}, can also be physiognomized, as can grammatical categories.[57]

---

[57] For example, there is a discussion of the how the ablaut (the German "strong verb" preterite marker) may be physiognomized by German subjects, because it feels like the modulation of a primary vowel, "to depict the modification or modulation of the action in going from a present status to a past status" (Werner and Kaplan, 1963: 236).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Many more insights regarding phonetic symbolism could be gleaned from their book, but the other essential point which theses authors have to make is on what they term the "plurisignificance of vehicular material" (Werner and Kaplan, 1963: 216). They describe the application of core meanings (such as the thinness and penetratingness of *i*, the roundness or hollowness of *o*, the mixed dark-light quality of the diphthong *au*, the greater articulatory tenseness and decreased purity or clarity of umlauted vowels) in the physiognomization of different words containing these vowels. A wide range of meanings can be thereby derived: all underlyingly related, and yet different—the phonetics highlighting different semantic aspects of the words in question, and the semantics bringing out different aspects of the phonetics. (Werner and Kaplan, 1963: 217).

Given that the original act of physiognomizing an utterance *already* assumes a translation from sound to other modalities—the authors point out that even the infant's onomatopoeic imitations involve translations into and out of her own developing phonology (see Werner and Kaplan, 1963: 102)—it is not surprising that the process of deriving and attributing meaning is characterized by enormous transferrability. It is this notion of plurisignificance (and the failure to understand it and its implications) that has led to so much confusion in phonetic symbolism experimentation. "The fact of plurisignificance of vehicular material has had a fundamental relevance to the much-discussed problem of 'sound symbolism.' Various studies have demonstrated that vowels, in particular, are correlated with certain meanings. Peculiarly enough, these findings, and the conclusions drawn from them, have tended to obscure rather than to promote and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

understanding of what occurs in 'matching sound patterns to meanings.'" (Werner and Kaplan, 1963: 218).

Sapir's and Newman's findings on correlations between vowel quality, size and brightness "led to expectations that languages such as English might make extensive use of this type of sound symbolism in the formation of its (sic) vocabulary." But Werner and Kaplan point out that phonetic analyses of the lexicon in the hopes of "finding such a universal relationship in a living language, especially an Indo-European one, derives, we believe, from two erroneous assumptions." The first is that "synesthetic links play a dominant role in the phonemic patterning of any contemporary language," which is "certainly at odds with the contention of most linguists, who maintain that onomatopoeia, based in part on synesthesia, is only of minor importance in the lexicon of a contemporary language: the dominant principle underlying word-formation, and changes in the external forms of words are of quite a different order, typically overriding any onomatopoeic tendencies.

"The second erroneous assumption may be termed the 'fallacy of constant elements,' that is, the fallacy which presumes that there are fixed 'bricks' from which verbal forms are built for the conveying of meanings. This elementaristic assumption enters into the interpretation of a certain vowel (or consonant) as entailing a fixed and constant signification, such as largeness or brightness. *As we have indicated, single sounds are plurisignificant; that is, they possess a wide range of dynamic properties. Furthermore, single sounds rarely determine the symbolic pregnance of a word: whether a sound has*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

this or that expressive value, or any expressive value at all in a word, depends on the dynamic structurization of the word as a whole. Sometimes these pervasive dynamics of the word as a whole are of a homogeneous simplicity, completely overriding the characteristic forms of the single elements... More often, there is a heterogeneous complexity, i.e., the various parts of a word contributing different features to the total physiognomy...

"Because these erroneous assumptions have furnished the theoretical ground and justification for most of the studies on 'sound symbolism'—and have thus oriented attention towards the presence (or absence) of manifest onomatopoeias in languages or towards the constancy (as manifested through consensuality) of the expressive values of vowels and consonants—the basic issues in the problem of the expressive representation of meanings through sound-patterns have either been overlooked or obscured" (Werner and Kaplan, 1963: 218—219, emphasis added). This has been quoted, apart from examples, in full as essential commentary and worth presenting in the authors' own words.

The authors refer to an unpublished doctoral dissertation by T. Iritani (1962), which centralizes the issue of how to bring out true phonosymbolic responses to artificial word stimuli, and not just to elicit a range of apparently random free associations. (This is the same issue addressed by Bentley and Varon (1933), who came to the conclusion that structuring the responses by asking for polar judgements was introducing a forcing, falsificatory element.) Iritani gave 30 subjects 35 artificial words, and asked them to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

associate to whatever was evoked by their sound patterns. He obtained, not surprisingly, a wide variety of responses, just as did Bentley and Varon and other experimenters when not limiting subjects by providing a "contrasting universe of discourse or stipulation of a dimension" (Werner and Kaplan, 1963: 220).

However, Iritani hypothesized that *underlyingly* the same phonetic features might be triggering the same semantic responses—even though the features would be interpreted differently on the surface by the different individual subjects.[58] So he analyzed all the referents given to his stimulus words in terms of six polarized dimensions, including small/large, bright/dark and round/angular. His findings "suggested that certain patterns 'sub-evoked' the same expressive property in many members of his sample—an expressive property masked by the subsequent elaboration of the experience into a complex referent" (Werner and Kaplan, 1963: 221).

A second experiment then restricted "the universe of discourse" by asking for judgements of artificial words along the six dimensions. More than 40% of words evoked a consensus. 31 out of 35 yielded consensual choice with regard to at least one dimension; 26 words out of 35 had multidimensional expressiveness. This experiment indicated that, as any poet could attest, a number of words very different in phonetic material "functioned in the experimental situation as expressive synonyms," and thus "patterns which are quite different from a geometrical-technical point of view may, through

[58] For description of sets of similar individual highly varied responses generated by the same stimuli and clearly having interpretation of the same features of the stimulus in common, see Osbert, Karwoski et al (1942).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

physiognomization, be apprehended as quite similar to each other..." (Werner and Kaplan, 1963: 225). Thus, not only are words plurisignificant, but different words which possess "similar dynamic-expressive features" can be expressive synonyms, and those possessing opposite features, antonyms. (This factor may indeed account for some of the mysteriously successful matches in the word-matching studies to be discussed in Chapter 5.)

The question of antonyms and oppositions is extremely important; it comes up again and again in phonetic symbolism literature. In a later chapter of *Symbol Formation*, discussing experiments which matched line drawings to referents, subject consensus was seen as achievable despite the issue of plurisignificance—but it depended on several conditions. One was *context*; there must be an orienting shared "universe of discourse." Another was a *clear degree of differentiation between stimuli.* Yet another was that the stimuli (drawn, in this case, but the principle would also hold for artificially composed words) *be depictively incisive*: that they should effectively express referents by easily lending themselves to physiognomization. However, for our purposes, the most important requirement was the *degree of differentiation between referents*: "only when referents can be sufficiently distinguished—sharply polarized—within a common context, is consensus with regard to the representations of those referents likely to occur" (Werner and Kaplan, 1963: 349).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This experimental requirement must be born in mind when considering certain issues raised by the word-matching studies, specifically the debate over the need for antonyms in order to get good matches. (The use of minimal pair stimuli acknowledges the need for just this kind of contrastive polarization of meaning, as in Sapir's (1927) original *mal* and *mil* used to distinguish a small from a large table.) But the need for antonyms also must be linked back to Darwin's second great principle of gestural communication, that of Antithesis, and to the Saussurean discovery of oppositions as being the central carrier of linguistic meaning (see p's 12-13 of this dissertation). To investigate phonetic symbolism is to put oneself into a situation where meanings may indeed be intuitively communicated and communicated with enough consensus for the researcher to say that a certain sound or combination of sounds "signifies" this quality or referent. However, in this linguistic area there is no commonly acknowledged authority to legislate meaning—no dictionary or grammar book. Hence the necessity for polarizing referents: it is a way of immediately orienting the subject in a semantic terrain which has no apparent signposts.

In 1963, Oyama and Haga undertook a study to "analyze the nature of 'intrinsic factors' ...considered to intermediate between the figural and phonetic symbolism" (Oyama and Haga 1963: 132) Once again, the instrument of measurement was the Semantic Differential. Three separate experiments were conducted. The first was to determine attributions to 14 figures, which included a square, a circle, two triangles and figures taken from Köhler's (1929) and Fox's (1935) work. The second investigation was of attributions to 16 trisyllabic[59] nonsense words, while the third, performed two years later,

[59]This was a Japanese experiment; even though involving less variables, monosyllable stimuli would have

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

matched the visual and verbal stimuli from the previous two experiments, both directly by subjects and indirectly by comparison with their profiles obtained through the earlier tests using the semantic differential.

In the first experiment, 40 female subjects judged the figures on 35 scales, with order of presentation of both scales and figures carefully controlled. Via factor analysis, five factors were extracted from the scales used; potency, evaluation and activity accounted for respectively slightly over a third, a quarter and a tenth of the total variance.

Thirty-nine new female subjects took part in the second experiment. Artificial word stimuli included Köhler's *baruma* (modified for Japanese subjects) and *takete*, and a minimal quintet of words: *ramara, rimiri, rumuru, remere, romoro*. Presentation was both aural (one pronunciation of each word) and visual (in a booklet). Striking Potency findings on the vowels in the minimal quintet were "that *u* and *o* were conceived deeper, farther, fuller, softer, heavier, hotter, wetter, and more smooth, stable and unreal than vowels *i* and *e*, and that vowel *a* had median values" (Oyama and Haga 1963: 136). In terms of Evaluation, "*i* was rated most highly on all scales... in the directions of clear, fresh, new, sharp, distinct, intellectual, light, fast, tense and small, in a sharp contrast with the vowel *o* which gave its lowest value on all of these scales in the direction of muddy, stale, old, blunt, vague, unintellectual, dark, slow, relaxed, large and unhealthy" (Oyama and Haga 1963: 132), with profiles of the other three vowels located between the extremes of *i* and *o*. Activity scales indicated no clear tendencies, "except that *a* was felt

---

been perceived, presumably, as highly unusual by the subjects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

strong and manly, *u* womanly and *o* negative" (Oyama and Haga 1963: 136)."The vowel *o* was also judged as the most boring, static and commonest sound" (Oyama and Haga 1963: 137).[60]

Subjects in the third experiment were 102 new females. On the direct matching task, it was found that curved forms were felt appropriate to words with *m*, *r* or *a* and straight-line forms to words containing *k*, *t*, or *i*. Matching of the *baruma/takete* items was in agreement with the findings of Köhler and Fox, but other, similarly contrasted artificial words acheived greater fittingness scores. When comparing the direct matching and the semantic differential evaluations, the experimenters found "fair" but not "complete agreement" (Oyama and Haga 1963: 141). They concluded that the Potency factor was mostly responsible for attributions of similarity between figural and phonetic symbolism: "perceived angularity, sharpness, hardness, distinctness, tension and strength are considered to be most dominant in their effects" (Oyama and Haga 1963: 142).

In 1964, Holland and Wertheimer ran two experiments to determine the triggering factors leading to fittingness judgements. They were looking for "physiognomic similarity" as a possible basis for understanding "some of the functional determinants in naming" (Holland and Wertheimer 1964: 111). Their first experiment used Köhler's *maluma*[61]*/takete* quartet of words and figures with 148 subjects who were asked to judge

---

[60] See Hamano (1986: 161-162) who asserts that *o* is "modest"—a low-profile vowel.

[61] This was actually *baluma* in the first experiment, and then (to avoid associations with *balloon*) *maluma* in the second.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

both figures and words separately on 10 different scales from the Semantic Differential. The conventional pairing of figures and words were found to be reflected in similar semantic ratings on the scales.

Additionally, on more detailed analysis of results (by comparing mean distances between scale values for *takete* and those for the angular figure, and scale values for *baluma* and those for the rounded figure) the experimenters found that only 3 out of the 20 scales did not show statistically significant closeness of values between fitting verbal and visual stimuli. This was the case whether the scales were 'literal' like angular/rounded or not, "showing that physiognomic similarity within a fitting pair is not limited to simple descriptive characteristics" (Holland and Wertheimer 1964: 113).

A second experiment took the same two figures and the *maluma/takete* pair, but added three more items to the verbal stimuli: the letters *U* and *K*, found in pilot studies (undescribed) to have "physiognomic properties closely resembling those of the entire names" (Holland and Wertheimer 1964: 114), and the composite *kelu*, having elements of both names. The 28 subjects were given 5 semantic differential scales, plus a supplementary scale indicating direct judgement of fit. On the direct judgement scale, the finding "strongly confirmed Köhler's original observations... at least with the present stimuli, the fittingness phenomenon in naming does occur" (Holland and Wertheimer 1964: 114). The letters *K* and *U* were found to be "strikingly similar to their corresponding full nonsense words" (Holland and Wertheimer 1964: 115), and this was also true of those letters on semantic differential judgements. *Kelu* "fell as would be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

expected about the middle" of the direct judgement scale (Holland and Wertheimer 1964: 115), but on semantic differential ratings, it was judged as significantly more similar to the rounded figure.

The authors concluded that "the physiognomy of the nonsense words may reside to a great extent in their component letters" (Holland and Wertheimer 1964: 117), but their findings bring up a couple of issues. Firstly, in the case of *K* and *U*, there was a sound-grapheme confusion, and thus a real weakness in the experimental design. What exactly was being evaluated: two letters, two sounds, or the contrast between a vowel and a (plosive) consonant? Secondly, the semantic differential judgements on the composite *kelu* seem to show clear signs of the domination of one syllable element, *lu*, over the other, *ke*. Several different variables may have been responsible for this: the position of each syllable in the word, dominance of the vowels, dominance of the consonants, feature-contrasts of vowel-length, consonantal liquidity versus plosiveness, and so on. Interesting but enigmatic results such as this call into question the use of any but the most limited (and systematically contrasted)[62] word-stimuli in order to evoke pure, unconflicted evaluations.

Also in 1964, Quenk published a short dissertation on fittingness of words to objects. He used 12 spheres contrasting in size, texture (dull versus shiny) and color (yellow or orange versus brown or black). His subjects were all native English speakers, selected as having little or no experience of foreign languages. Only 23 subjects took part, a control

[62] In other words, minimal pairs were used.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

group of 11 and an experimental group of 12.[63] Both groups were asked to construct as many designs as they could within the experimental hour, using all the spheres (which were connected by sticking dowels into them) and describing aloud each step of the construction. In recorded instructions, the control group were given 'fitting' names for the spheres: the big ones were *yuf* and *gup*, while the small ones were *feep* and *steeg*. The experimental group were given 'nonfitting' names for the spheres, i.e., the names were assigned in reverse. Each design was timed, and the number of times the spheres were named was noted. Then subjects' first and last pronunciation of all four sphere names was edited out of the recording of the task session, and the names were spectrographically analyzed.

Results showed a definite and statistically significant shift in F_2 (i.e., in anteriority) in the experimental group's pronunciation of nonfitting names, moving them in the direction of fittingness. The control group's shift was found to be "negligible" (Quenk 1964: 33). The experimental group constructed fewer designs, learned the sphere names less well and named them less frequently. Even though these are interesting results, and Quenk's design was ingenious, his subject numbers were very small. Moreover, the phonetics of his experiment were dubious. He claimed that his stimuli opposed back and front vowels (Quenk 1964: 27-28), but the vowel in *yuf* and *gup* is often classified as central (see Ladefoged 1982). In addition, Quenk's vowels contrasted along a dimension (lax/nondiphthongized versus tense/diphthongized) which he did not apparently take into

---

[63] Presumably, since the procedure involved one-by-one observation of subjects, taping and transcription, a larger pool would have been impracticable.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

account. Consequently, this may have resulted in a fittingness conflict, with the *longer* vowel (Peterson and Lehiste 1960) denoting the *smaller* spheres.

After two previous studies matching artificial words to figures which had produced very inconsistent results, Weiss (1964b) hypothesized that *meaningful stimuli* were necessary "to significantly elevate agreement scores" (Weiss 1964b: 256). He therefore designed an experiment in which four sets of scores could be compared in terms of referent or stimulus meaningfulness. These scores were obtained through evaluations of matching of two groups of referents (either visual or named) with two groups of verbal stimuli Verbal stimuli were bisyllabic CVCV artificial words "pretested for meaningfulness by two groups of judges" (Weiss 1964b: 256). Unfortunately, Weiss did not describe the test, or how he arrived at meaningfulness criteria.[64] Possibilities were as follows: a) low meaningfulness artificial words to be matched with non-representational figures, b) low meaningfulness artificial words to be matched with "verbally presented" (Weiss 1964b: 256) meaningful objects (such as a raincoat, tin can, hatchet or horse), c) high meaningfulness artificial words to be matched with non-representational figures and d) high meaningfulness artificial words to be matched with verbally presented meaningful objects.

Subjects were 88 female students, divided into 4 groups. Their tasks were to match 30 pairs of words either to figures or to named objects along 3 dimensions: magnitude,

---

[64] Examples of "high meaningfulness words" were *bubu-mumu* or *bobo-popo*, and of "low meaningfulness words," *bebe-meme* or *fafa-vava* (Weiss 1964b: 257).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

brightness and sharpness. Verbal material was presented on tape, and visual material with a projector. Using named objects (i.e., meaningful referents), significant agreement was reached on brightness. There was consensus on size but without reaching significance, while angularity scores were not affected by using named objects. The high or low meaningfulness of the artificial words apparently had no effect on scores.

Weiss suggested that studies of phonetic symbolism needed "a meaningful referent in the presence of which the judgement is to be performed... availability of a dimension[65] may be necessary but is not entirely sufficient. The subject... must also have a stable or defined referent to which he can relate the dimensions before results consistent within a group can be expected. *It is our belief that, simply from listening to a pair of contrasted sounds, a subject has no way of 'knowing' which of the pair is 'larger,' 'brighter' or what have you*" (Weiss 1964b: 260-261, emphasis added). In this study, Weiss was taking a further step in the discussion of forced choice in phonetic symbolism experiments. The reader will note that in the italicized quote above, Weiss referred to "sounds," but in fact the issue was more complicated, since he was actually discussing attributions to *words*: "When asked to make such judgements (i.e., of magnitude or brilliance), subjects tend to associate to the stimulus words in order to come up with some frame of reference.... (they) try to make use of the totality of their sense experiences which might bear upon the judgement in hand... Where the approach to the problem is prescribed in such a manner

---

[65] Examples would be dimensions such as size or brightness.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that the associative chain leading to the response is subject to greater control, higher agreement is likely" (Weiss 1964b: 261).

Weiss' original inspiration was Sapir's (1927, 1929) use of both the magnitude dimension and the referent 'table' in obtaining phonosymbolic judgements (Weiss 1964b: 255). However, he also raised the question, "How is it that word-matching studies which make use only of a dimension in the form of an English antonym pair [cf. Brown et al] achieve positive results?" (Weiss 1964b: 261). Two possible answers were the presence of "actual similarities in the structures of words" to be matched from the different languages, and "that a group of subjects with essentially similar backgrounds and comparable intelllectual makeup may share a sufficiently similar pattern of associations to allow for some agreement, even in the absence of a specific referent" (Weiss 1964b: 261).

It will be seen that Weiss' second explanation leans towards culturally mediated experience and thus learning theory rather than "the *gestalt* position which is based upon the concept of physiognomic perception" (Weiss 1964b: 262). Yet he also remarked that in his own investigations, the discrepancy between Brightness and Magnitude judgements (i.e., greater versus lesser consensus) indicated that shared cultural-linguistic experience could not account for "the whole story" (Weiss 1964b: 262).

Two problems are apparent in this study. The first is, as mentioned above, Weiss' lack of reported criteria or experimental procedures for establishing the high meaningfulness versus low meaningfulness of his artificial word stimuli. The second is that unfortunately

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the referents of the stimulus words were contrasted in *two* ways: both visually (drawn figures) versus orally (words such as 'horse' or 'hatchet') *and* in terms of their meaningfulness versus meaninglessness. Thus, subjects may have been responding to the difference in media as well as—or instead of—the difference in meaningfulness. It is regrettable that the experimental design did not include *figures* of horses and hatchets, thus excluding the unnecessary variable.

In 1964, Langer and Rosenberg made an evaluative study matching semantically meaningful artificial words with subject evaluations on antonymic semantic scales. The artificial words[66] were created by "speakers" as representative of two dimensions, complex space and color. They were then to be assessed by "listeners" first along three and then five dimensions: simple space, complex space, color, moral evaluation and direction.

The main experimental question was how well such artificial word stimuli can communicate meanings; this was broken down into three sub-questions. Firstly, do listeners evaluating along the *same* dimension (e.g., color, for color words) agree significantly with the meanings to those words originally assigned by speakers? Secondly, when listeners are evaluating stimuli, for example color words, along *different* dimensions, such as morality or direction, do the originally-generated meanings affect judgements along those quite other dimensions? Finally, do some artificial words fit some dimensions better than others?

---

[66] These were referred to confusingly throughout their report not as *words* but as *sounds*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Two investigations were undertaken, the first being evaluation of stimuli representing complex space. Fourteen speakers created monosyllables to symbolize four spatial categories: "tall and wide," "short and wide," "tall and narrow" and "short and narrow." (The authors gave no reasons for complicating their (listener) subjects' evaluational task from the outset by demanding *two* semantic dimensions from their speakers.) Twentyfive words representing all four categories were then randomly selected and were recorded by a speaker unaware of their "meanings," thus ruling out the chance of an expressive performance. The authors then presented the stimuli simultaneously visually and orally to 75 subjects who were to evaluate them along three semantic dimensions: simple spatial, color and complex spatial.

In the second experiment, 17 (new) speakers created monosyllables to symbolize "red," "green," "yellow" and "blue." Twentyfive of the resulting artificial words were then recorded by the same unaware speaker and presented to 119 new subjects for evaluation along the same 3 dimensions as above, plus morality (goodness) and direction. Results from both experiments demonstrated above chance agreement where speakers' original meanings could be matched with listeners' judgements along the same dimension. Furthermore, a majority of words evoked consensus[67] along the three dimensions which were not part of the original reference: simple space, direction and morality. The authors concluded that although artificial words can represent different referent dimensions, some

[67] However, the authors were somewhat vague as to how much of subject agreement was statistically significant. Consensus was classified as "overwhelming," or "predominant, but not overwhelming" (Langer and Rosenberg 1964: 368).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

referents are more easily designated than others. Simple space was the most often agreed-on, "whereas colors and evaluations are least appropriately represented, and directions occupy an intermediate position" (Langer and Rosenberg 1964: 370). Unusually for a phonetic symbolism experiment, they cited Werner and Kaplan (1963) in support for their view that "some referent dimensions may be metaphorically expressed and consistently understood in one symbolic code but not in another" (Langer and Rosenberg 1964: 370). However, the underlying question of which segment or segment combinations might represent these dimensions could not be answered by their investigation.

Apart from the initial complication of having created stimuli that represented simultaneously two referent dimensions, there was another problem with this study, admitted in a concluding footnote. The authors pointed out that the relatively uncontrolled nature of the stimuli (monosyllabic, but using any combination of segments) meant that their conclusions must be tentative: "Any firm conclusion must wait upon an enquiry in which sound patterns are systematically varied in terms of beginning and ending consonants, and beginning, middle and ending vowels" (Langer and Rosenberg 1964: 370).

In a study using simple, controlled stimuli Shtern (1967) asked 30 Russian subjects to arrange 6 single Russian vowels "in order from highest to lowest (pitch). *None of the subjects doubted that the problem had a solution*" (Shtern 1967: 17, emphasis added).[68]

---

[68] The reader is reminded of von Hornbostel's remark about asking a subject to find on the piano a tone equivalent in brightness to the smell of lilac: "generally he thinks the task nonsense at first, but if he can be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This resulted in the following ranking: *i, a, o, e, u, y [bl]*. The endpoint vowels, *i* and *bl*, yielded majority reponses of respectively 21 and 18.[69] In a second experiment, 90 subjects were given 12 colored discs representing 6 colors and 2 degrees of saturation. Their task was to assign the same 6 vowels to the colors. Shtern classified his subjects into four groups: a) a small number of "synesthetics, persons with color hearing," who "confidently solved the problem in 30-60 seconds. It seemed as if they had already encountered the problem...often... we heard these admissions: 'It has seemed that way since I was a child.'" b) subjects forming phonisms, who "immediately took the discs in their hand and said.. 'This color is *a*'" c) subjects forming photisms (the largest group), who "pronounced the vowels aloud and selected a color while listening to the sound," and d) "subjects who solved the problem logically, i.e., by reference to some common property. For example, ' *i* is very sharp and piercing. Hence it is dark green.'" (Shtern 1967: 18).

In "a preliminary analysis of the data within each group" Shtern discovered, most strikingly, that there were "no differences among the results of different groups" (Shtern 1967: 18), a finding similar to those of the experiments of Odbert et al. and Karwoski et al. (1942). However, when all the results collected in a general table, "data scatter was high. It was suggested that the decisive factor in color-sound associations was not the specific spectral characteristics but more general parameters—brightness and the warm and cold spectral regions" (Shtern 1967: 18).

---

persuaded to deal with such nonsense at all, it goes very well" (von Hornbostel 1927: 84).

[69] However, for these purposes, 30 would be a minimal sample.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

When the data were reanalyzed, the following noncolor evaluations were obtained according to "the statistical majority of the subjects... *a* – warm; *i* – cold; *o* – light;[70] *bl* – dark; *u* – dark and cold" (Shtern 1967: 18), while no parameters were obtained for *e*. Again, Shtern noted "the results were not influenced by the speciality of the subjects," concluding that "the majority of people associate speech sounds with nonaural sensations" (Shtern 1967: 18).

[70] This recalls Grammont's (1946: 384) classification of *o* as "*éclatant*" (brilliant).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 5

PHONETIC SYMBOLISM: TWO LINES OF PSYCHOLINGUISTIC EXPERIMENT FROM 1934 TO THE PRESENT

This chapter focuses on the two most important types of phonosymbolic experiments: word-matching studies, whose design originates with Tsuru and Fries (1933), and studies using a range of minimal pair stimuli, originating with Sapir (1927, 1929). The experiment described in the later chapters of this dissertation belongs to this second type.

We begin with studies of word-matching, for a number of reasons. Spanning the period 1955-1971, they form a chronologically more tightly related group than do minimal-pair experiments. Also, unlike minimal pairing, a technique still relevant in the search for phonetic symbolism, the word-matching approach is now better seen as part of the history of linguistics than as having developed a useful method of getting phonosymbolic data. The word-matching experiments were an attempt simultaneously to demonstrate cross-linguistic occurrence of (in Taylor and Taylor's (1965) terminology) objective *and* subjective phonetic symbolism, i.e., to find phonosymbolisms in the lexica of more than one language and to show that interlinguistically subjects are sensitive to them. The quest for universal linguistic structures was renewed in the Fifties,[1] and it is no coincidence that

[1] Chomsky's *Syntactic Structures* was published in 1957.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

at this time we find a great interest in establishing or disproving the existence of universals in phonetic symbolism. Finally, consideration of the arguments, insights, errors and failures associated with this group of experiments will allow us better to appreciate the economy and efficiency of minimal pair investigations, and the resulting solidity of many of their findings.

5.1 Word-matching experiments

Brown, Black and Horowitz (1955) designed a word-matching experiment built upon the reports of three unpublished studies: Tsuru (1934), Allport (1935) and Rich (1953).[2] The aim of all these experiments was to discover whether subjects would correctly pair up words with the same meanings, either in English and an unknown language, or in two unknown languages. Brown, Black and Horowitz selected twenty-one English antonymic pairs naming sense experiences exclusive of aural ones (e.g., *heavy-light*) and of fairly frequent use; the list of English pairs were then translated into Chinese. Czech and Hindi.

---

[2] Tsuru presented a list of thirty-six Japanese antonym pairs for correct matching with their English equivalents to fifty-seven non Japanese-speaking subjects. Each pair was presented in two modalities, visually, in Romish characters and aurally twice, in both possible orders; 69% of the words were correctly guessed. However, Brown, Black and Horowitz cautioned that the English translations of the Japanese words could have been phonetically influenced by their Japanese equivalents or vice versa, since Tsuru made his selection with full knowledge of both languages, and translator bias could have resulted in stimuli especially easy to guess. Allport improved on Tsuru's design by having Tsuru's list translated by a (presumably Japanese-speaking) Hungarian without reference to the English list, but this still did not rule out the possibility of the Japanese words' phonetic structure influencing the Hungarian translations, since any time that words are translated, the phonetics of the original item may influence the choice of word in the target language. Allport's sixty-eight subjects guessed correctly 56.6% of the time, "still above the chance level" (Brown, Black and Horowitz 1955: 389). Rich had forty-four subjects match twentyfive word pairs in English, Japanese and Polish; both the Japanese and Polish sets were matched above chance level, at respectively 57.2% and 64.8%.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Presentation to the eighty-six subjects was both visual and aural.[3] As there were no significant differences between the performances of the men and the women, data were combined; matchings for both Chinese and Hindi were approximately 59% correct, and for Czech approximately 54% correct. However, two aspects of these and all subsequent results must be considered: *agreement* between subjects and *correct guessing*.[4] Only correct guessing can be held to indicate cross-linguistic phonetic symbolism.

Checking for possible hitherto uncontrolled influences on guessing, the experimenters looked at the factor of different word-lengths within a pair, but found no indication that this affected subjects' performance.[5] Another possible influence, the expressive quality of a speaker's voice, was also found not to have had an effect.[6] Results were considered as due either to some residue of "universal, unlearned intersensory connections still present in the vocabularies of the four languages tested," or to the subjects' common experience of being native English speakers "who naturally share a cultural conception of the sounds most likely to express a given meaning" (Brown, Black and Horowitz 1955: 392-393). The fact that 25% of the reponses were wrong, but showed significant

---

[3] Recordings of the foreign words were played, while subjects were given a page on which English pairs on one side corresponded to a foreign language pair on the other. Subjects were allowed ten seconds to match the English pair with their foreign equivalents.

[4] The authors calculated that 64% of judgements had to be in agreement to reach the .01 level of significance. Out of each list of twentyone pairs, nine or more items were agreed-on and correct in each language, and four or five agreed-on but incorrect. They concluded their subjects' conceptions of the proper translations were right twice as often as they were wrong.

[5] It was noted that performance on (monosyllabic) Chinese words was superior to Czech, where word-lengths varied.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

agreement tended to support the second explanation.[7] The experimenters concluded their technique had reliably unearthed evidence of universal phonetic symbolism, ascribed to gestural roots.[8]

Maltzman, Morrisett and Brooks (1956) then challenged Brown et al.'s claim in their report of another word-matching experiment using Japanese, Croatian and English. Their stimuli were presented in the visual mode only;[9] one stimulus word was to be matched with one of a pair of response words, and Tsuru and Fries' (1933) word list was used. Subjects were sixty-five undergraduate and graduate students in psychology. Their first experiment investigated the effects of order and type (i.e., single word or pair) of presentation of Japanese and English. Since it was found to be immaterial whether a Japanese single word was matched with an English pair or vice versa, the resulting data were combined; findings of significant correct guessing were obtained. Then, in order to correct the possible partiality of Tsuru's translation, the word list was retranslated into

---

[6] This was checked via a control group (number of subjects unspecified) that responded only to the printed sheets without the aural component. Their overall performance was slightly more successful than that of the experimental group.

[7] However, if those "cultural conceptions" were based on sound-meaning connnections rooted in the lexicon of the speakers, then it might be expected that more correct translations would have been made from related than from unrelated languages. This was not the case. It was noted both in Brown, Black and Horowitz' study and in the experiments of Tsuru, Allport and Rich that performance on non Indo-European languages was no worse than performance on Indo-European ones.

[8] "We are inclined to believe... most subjects responded in terms of principles of phonetic symbolism imperfectly realized in all languages... our results (are) indicative of a primitive phonetic symbolism *deriving from the origin of speech in some kind of imitative or physiognomic linkage of sounds and meanings*" (Brown, Black and Horowitz 1955: 393, emphasis added). For the origins of this theory, see Darwin (1955) and Paget (1930).

[9] Subjects were explicitly instructed to avoid systematizing responses.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Croatian and given to forty-eight subjects for matching with English, and the same level of correct agreement was obtained

However, the authors wished to put hypotheses of universal phonetic symbolism to a more stringent test, by using stimulus sets in which there were no English words. They found that when non-English stimuli alone were used, there were no significant results, and concluded that for matching to be successful by English-speaking subjects, one member of the match had to be an English word. When the findings of Maltzman et al.'s are added to those of Brown et al.'s (i.e., of approximately 75% agreement, with one third of that total's being *wrong* agreement), it appears that what was being tested in these experiments was the phonetic symbolism of English projected onto the vocabularies of other languages.

A report on yet another word-matching experiment appeared the following year (Brackbill and Little 1957). These experimenters were also sceptical about Brown Black and Horowitz' (1955) methods and findings, and determined to make a properly rigorous investigation. They had four criticisms of Brown et al's design. Firstly, visual presentation was inappropriate for testing phonetic symbolism, a phenomenon they considered dependent on auditory cues.[10] Secondly, the fidelity of Brown et al.'s foreign translations was not checked;[11] a consensus of translators, and a further step of back-

---

[10] Also, to avoid the influence of private, idiosyncratic associations on the judgements of English-speaking subjects, Brackbill and Little claimed that English words should not be used as stimuli: matching should be done only with foreign-foreign pairs. (However, they did not follow their own guidelines.)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

translation should have been provided. Thirdly, they found Brown et al.'s "population of concepts... indeterminant and the sampling procedures...ambiguous" (Brackbill and Little 1957: 312).[12] Finally, an investigation of a supposedly universal phenomenon should have used a more representative sample of the world's languages than Brown, Black and Horowitz' (where three out of the four languagese used were Indo-European). Maltzman, Morrisett and Brooks' (1956) study was criticized as having similar methodological drawbacks.

Brackbill and Little's own experiment was designed carefully to examine three aspects of of the word-matching process. These were the influence of modes of presentation on guessing accuracy (visual, audio-visual, aural), the degree of accuracy obtained when the word sample was a random one from a determinate population, and determination of specific stimulus characteristics leading to accuracy. Four non-cognate languages were tested, Hebrew, Chinese, Japanese and English, and the word-population was chosen from among the five hundred most frequently-used items. In order to obtain reliable non-English stimuli, a meticuluous process of translation and selection was undertaken, ruling out (as far as possible) any translation bias or subjective element of stimulus choice.

The authors had already established in a separate pilot study (no details given) that subjects could guess meanings as accurately whether stimulus words were presented

---

[11] Concepts might exist in one language but not in another; thus, the translator could make an error or choose an English word which corresponded phonetically rather than semantically to the foreign original.

[12] For example, stimulus words did not all, as claimed, "name sense experiences," or have the frequency of "a hundred or over per million." Furthermore, some word pairs were not antonymic (e.g., *iron-gold*).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

singly or in pairs; therefore, in the interests of economy, single items were to be matched with each other via judgments of sameness or difference. [13] All three modes of presentation were used, each mode being used with sets of two stimulus types: English-foreign and foreign-foreign.[14] The 269 subjects were divided into groups of around forty, and subjects were given three sets of fifty word pairs corresponding to the three modes of presentation.

Results indicated an uncertain answer to the authors' first question about mode of presentation infuence; with English-foreign stimuli, visual presentation increased correct guessing, while with foreign-foreign pairs, the effect of visual presentation depended on which of the two languages had been paired. Findings on the guessability of random word samples from a determinate population also depended on the language pairs used.[15] In analyzing which stimulus cues had favored correct guessing, Brackbill and Little examined objective stimulus characteristics of the word pairs leading to high agreement (correct or incorrect); they found high-agreement items to comprise around 40% of any list. Three of the contributing factors discovered were phonetic: similarities in word length, in vowels (front vs. back articulation) and in consonants (manner). A fourth, similarities in hyphenation or spacing, was graphic, while a fifth factor was lexical: if words or word parts suggested some English concept, connotative associations might have led to similarity judgements.

---

[13] Subjects were also instructed that in only 50% of the pairs did the two words mean the same thing.

[14] This resulted in a total of eighteen presentation possibilties.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The stimuli were then reanalyzed,[16] and the five above-mentioned factors ranked in terms of their effectiveness as follows: connotation, vowels, length, spacing and consonants. However, since the authors still considered their analysis might have been over-subjective, a test using nonsense syllables was devised to check on the effects of the first three ranked factors. Forty-nine subjects were asked to match a stimulus word with the closest in meaning of a pair of response words, and the influence of each factor and the possible effects of combining the factors were methodically explored; however, presentation was visual only.[17] All percentages of agreement were significantly above chance, except when the effect of connotation was pitted against the combined effects of length and vowel similarity. When nonsense stimuli were presented visually, the strongest influence appeared to be lexical associations, followed by the most general appearance of the word, i.e., its length, while vowel anteriority had the least effect. But Brackbill and Little's methods seem to have been dubious in several respects.[18]

---

[15] Chinese-Japanese proved the most guessable, Chinese-Hebrew the least.

[16] The authors took into account such things as individual differences in the frequency of factors and the influence they exerted depending on stimulus type (e.g., connotation was stronger for English-foreign than for foreign-foreign) and correctness of consistent agreement.

[17] We note that here the authors abandoned their focus on the effect of mode of presentation and fell into the same trap of concentrating on the visual and neglecting the aural that they initially criticized other researchers for doing.

[18] For example in using visual stimuli, researchers showed the same confusion between *letters* and *sounds* as found in other experiments in phonetic symbolism which concentrated on psychology at the expense of phonetics. Thus, several of the stimulus nonsense words contained two contiguous vowel letters; to the normal English-speaking reader, this suggests diphthongs or long vowels, and not the same vowel sounds that were in the word to be matched with them. Moreover, several nonsense words were polysyllabic, which again meant (for English-speaking readers, naturally interpreting the words via English stress rules) that a sixth vowel, [ə], would have been added to the five that the experimenters thought they were presenting.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In discussion of their results, Brackbill and Little noted significant discrepancies between their findings and those of Brown et al. (1955) and Maltzman et al. (1956). Subjects in the two earlier experiments were able to guess the English meanings of Chinese and Japanese words significantly above chance, while their own subjects could not do this.[19] Positive performance on guessing with certain foreign-foreign pairs in both Maltzman et al.'s (1956) experiment and their own were ascribed to noncognate languages' possibly "employing similar patterns of sounds and length to designate similar meanings—at least among frequently used words" (Brackbill and Little 1957: 318). However, their general conclusions were against the hypothesis of a universal phonetic symbolism.

Following this, Brown and Nuttall (1959) then reattacked the issue with another word-matching study. This was designed to answer criticisms leveled at the earlier, Brown, Black and Horowitz (1955) study, and to explore the reasons for the different findings of Maltzman, Morrisett and Brooks (1956) and Brackbill and Little (1957), notably their failure to obtain positive matchings on foreign-foreign word pairs. Brown and Nuttall (1959) extensively criticised Brackbill and Little's (1957) selection and potential pairing of stimulus words.[20] They also noted that despite Brackbill and Little's dicovery of

---

[19] Mode of presentation was ruled out as a potential cause (since among the three modes used was the audio-visual one employed by the earlier experimenters), as was their single-word method of stimulus presentation, since they had already proved to their satisfaction in a pilot that single or double presentations made no difference to subjects' responses. They were left with differences in word-samples used, or possibly some interaction between this factor and the single-word presentation method.

[20] Brackbill and Little's study was criticised for including function words like *when, first, this* and so on, in their eagerness to employ a random sample of high-frequency items. Subjects had been asked simply to judge if single words were the same or different; the experimenters had paired 50% of the words on their lists with equivalents in another language, creating stimuli for "same" matchings, but the residue of words thus left for "different" matchings, were not necessarily antonymic or even contrasting. Consequently, half of the words offered for judging in the foreign-foreign pairs were without a semantic referent dimension of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

significant interaction between presentation method and language pairs, totals taken across language pairs for the three methods did not vary by more than about 1%. While acknowledging Brackbill and Little's criticism of Brown, Black and Horowitz' (1955) word selection, Brown and Nuttall (1959) felt it was still useful to retain the original list.[21]

Brown and Nuttal challenged Brackbill and Little's (1957) finding that it did not matter whether the selections from the languages to be matched were in the form of pairs or single words.[22] Accordingly, words were presented in three ways: in English-foreign pairs, in Brackbill and Little's "same-different" manner (i.e., one word matched with one word), and, following Maltzman et al., foreign-foreign pairs.[23] The languages tested were Hindi and Chinese; subjects were two large classes of students (numbers unspecified). All presentations were audio-visual. Results indicated that judgements on the English-foreign presentation were significantly better than with either of the other two methods. There was no significant difference between findings for the other two presentation methods, and language differences for these two methods were not significant either: "Evidently

---

any kind; they might be highly abstract in nature and very difficult to characterize in terms of perceptual symbolism. Moreover, these words were presented singly: *this* can be contrasted with *that*, or *first* with *last*, but Brackbill and Little's subjects had no helpful orienting opposition.

[21] The old list had yielded a sizeable phonetic symbolism effect when presented in the original manner, and Brown and Nuttall wished to see whether changes of experimental procedure woruld alter this.

[22] This "ought not to occur if there is anything in the phonetic symbolism hypothesis and which the present investigators judged would not occur with the original list" (Brown and Nuttall: 1959: 443).

[23]The English-foreign presentation reduplicated the original Brown et al. experiment. The foreign-foreign presentation was of a Chinese pair and an equivalent Hindi pair, the presentation of languages randomly ordered. It differed from Maltzman et al.'s procedure by giving a pair of stimulus words rather than a single one. For the same-different presentation, Brackbill and Little's prcedure was followed, with 50% of the two word-lists the "same," leaving a residue of "different" words, not necessarily antonyms.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the phenomenon is highly sensitive to the method of presentation" (Brown and Nuttall: 1959: 443).

Results were explained in reference to Sapir's (1929) and Newman's (1933) original findings, and to Wisseman's[24] on polysyllabicity. Brown and Nuttall postulated a set of internalized rules of phonosymbolic perception based on phonetic features of the words' segmental composition and on the number of syllables in the word; they reasoned that only the English-foreign pairs procedure allowed subjects fully to use their knowledge of these rules. If subjects were given an English word pair, they would know the referent semantic dimension to be used, but foreign antonyms might use any one of a number of different phonetic features to symbolize the opposition.[25] Thus, "if two foreign pairs were combined, the subject might fail to identify the referent dimension and so match words on the basis of particular phonetic resemblances" (Brown and Nuttall: 1959: 444). The authors also speculated that among possible referent dimensions, perhaps only size and its traditional associates, such as brightness, weight and speed, might function as phonosymbolic directives.

Clearly, a central issue in the use of word-matching techniques is whether explicit contrasts are needed to obtain correct guessing. McMurray (1960) was the first to

---

[24] For details of this reference, see Brown and Nuttall (1959: 444).

[25] The authors here identify one of the main problems of much of phonetic symbolism research: overcomplexity of stimuli. They also comment that in Brackbill and Little's (1957) experiment, the fact that subjects ended up matching words on the basis of phonetic similarities may have been forced by the single-word method used, which deprived subjects of an orienting referent dimension.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

introduce antonymic pairs drawn from twelve scales of Osgood, Suci and Tannenbaum's (1957) Semantic Differential Test into this type of experiment. These were to be matched with fifteen antonym pairs in Chinese, Czech and Hindi (drawn from Brown, Black and Horowitz' (1955) word list). Subjects were seventy-six psychology students, and stimuli were presented visually. "Remarkable agreement" (McMurray 1960: 168) was found in matching words to scales, with no notable differences between the languages; subjects apparently could contrast words on many dimensions, regardless of (denotative) meanings. Statistical analysis of the data revealed clusterings of correlating judgments on several dimensions, indicating two super-dimensions: *bright-fast-high-light-active* vs. *dark-slow-low-heavy-passive*, and *small-weak-harmonious* vs. *large-strong-discordant*. McMurray observed that those stimuli using near-minimal pairs or ones with contrasting vowels were most successfully matched, concluding that the *bright* etc. judgments correlated with vowel height and anteriority, and *dark* etc. judgments with the opposite vowel features. He also suggested that *large* etc. judgments may have been triggered by plosiveness, "hard consonant sounds," but conceded that "vowels and consonants no doubt interact in a complex way" (McMurray 1960: 173). The interest of this study is in how antonymic contrasts along many dimensions brings out intersubject agreement, and in the search for phonetic contrasts as responsible; McMurray declined to speculate whether results indicate English-specific or universal phonetic symbolism.

Weiss (1963a) questioned the nessity of antonyms for correct matching, presenting his stimuli in two forms: these were *antonymic pair* condition, and *mixed* (non-antonymic) *pair* condition. Each condition included two language matches: English-Hindi and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

English-Chinese; words were taken from the original Brown et. al. (1955) list. Subjects were 394 undergraduates with equal numbers of males and females, divided into four groups, one for each condition/language match. Presentation was visual only; afterward, subjects evaluated the difficulty of matching on a four-point scale. All four types were found to be matched with better than chance accuracy, with no notable differences between them. When analyzed, uniformity of results was not found to be due either to irregularity of individual word-pair scores (i.e., no non-antonymic pairs scored freakishly high), nor to submerged antonyms in the non-antonymic stimuli. Difficulty ratings for antonymic and non-antonymic matchings were the same, but across all four groups males apparently found the experimental task easier and guessed slightly more correctly. Weiss concluded that antonyms were unnnecessary for correct guessing: meaningful context was all that was needed, and that this was provided by the inclusion of English words. Moreover, the use of word pairs for both languages gave subjects the advantage of four pieces of information with which to work. Finally, he observed "evidence of more than a single factor, [which could be]...associations...or structural similarities," in his subjects' introspections (Weiss 1963a: 546). This suggests that a range of triggers, depending on the individual subject's reaction, might have been responsible for the same evaluation.

Johnson, Suzuki and Olds (1964) compared the performance of deaf and hearing subjects in an attempt to establish how far phonosymbolic cues are purely auditory (and not kinesthetic or visual). Their experiment required subjects to match two-syllable artificial words which included all five vowel letters to fourteen non-antonymic English evaluative adjectives of varied lengths, e.g., *wise*, *lazy*. Construction of the artifical words was based

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

on the work of Demerse (1941),[26] as checked against the good-bad ratings in Jenkins Russell and Suci's (1958) semantic atlas, with Jenkins' findings deciding choice of initial letters and Demerse for the remaining portions of the words. Subjects were eighty-six hearing college students, thirty-six deaf, and thirty-six hearing adolecscents; all presentations were visual.

Results indicated 57% significant agreement on matchings for college students (8/14 pairs), but only 50% in the expected direction (i.e., with phonetically "bad" words matched with "bad" adjectives, and vice versa). For the deaf adolescents, the authors found that matching was no better than chance, while for the hearing adolescents, only 29% of *significant* agreement was in the expected direction, although 11/14 pairs were agreed-upon and 'correctly' matched. Thus, although the authors conclude that their findings support a) the notion of phonetic symbolism, and b) a deaf-hearing difference in ability to discern it, this does not appear to be born out by their results. Their construction of artificial words seems needlessly complex (two syllables, with a different vowel in each, three different consonants, i.e., a minimum of five possible triggers of symbolism), and the assumption that consonants can be consistently evaluated as good or bad appears very shakily based.[27]

---

[26] The authors refer to Barbara J Demerse' unpublished Master's thesis (University of Iowa): *The scaling of two-syllable adjectives in terms of their sound and of their connotation.* Certain consonants were judged as *euphonious/cacophonous*, others as *pleasant/unpleasant* and judgments of *pleasant* and *euphonious* correlated, as did their antonyms. For example, stops were "bad" (cacophonous, unpleasant) and nasals, "good" (euphonious, pleasant).

[27] For example, the Jenkins atlas is quoted as suggesting initial *c* begins "good" words and *g*, "bad" ones, i.e., two graphemes, but four phonemes are simultaneously indicated, yet the researchers see no problem with this (see Johnson, Suzuki and Olds 1964: 234).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In an article on the methodological problems in word-matching experiments, Barik (1964) criticized the visual presentation method. Visual presentation has two advantages: it is less easy for the subject to miss one item, as may happen in a sequence presented auditorily, and it bypasses inadvertently expressive use of the voice by the presenter. However, it may entail serious drawbacks. Firstly, it assumes that foreign spellings are always readable by the subject; secondly, even if the subject is successful at decoding the items, pronunciation is most likely to be filtered through an L1 phonology if no aural models are available. Thirdly, this method of presentation may put pressure on the subject to match words on visual bases: structural similarities (word-length, vowel and consonantal structures), connotational properties, semantic or syntactic clues.[28] Barik remarked that if visual presentation was used, items should be presented in a one-by-one sequence, so that there would be no chance of cross-item comparison (another drawback to this method). Finally, if part of the basis of phonosymbolic judgment was proprioperceptive (as implied by Sapir's (1929) and Newman's (1933) analyses) and if stimuli were in a foreign language, the subjects' imitation of a native-speaking aural/oral model would be essential, to minimize L1 phonological interference. Yet, paradoxically this would mean that auditory presentation alone would not be sufficient. Barik concluded that for subjects to attempt imitating non-native sounds, presentation should probably include those visual clues to articulation which were provided by observing the speaker.

---

[28] For example, attention to the distribution of suffixes in Weiss' (1964) Japanese stimuli would reveal which were verbs and which adjectives, and thus allow correct matching with the English verbs and adjectives provided.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Presentation by auditory (oral) method alone was used in an investigation by Atzet and Gerard (1965). Subjects were sixty Navajo Indian students (aged fifteen to twenty), with no knowledge of Chinese or Hindi, the languages used for matching. Twenty antonym pairs naming sense experience, drawn from Brown, Black and Horowitz' (1955) list, were translated into Navajo, using the commonest forms of those words. The matching task was given in the form of a pair of Navajo antonyms (AB) followed by a pair of Hindi or Chinese words in two orders (CD; DC). Each of the Navajo antonyms was matched with the foreign pair in both orders;[29] thus, there were four choices per foreign language, resulting in eight for each Navajo pair. Presentation was randomized, and presentation variation was controlled for, as was subject fatigue (the total task took two hours).

Results indicated guessing at no better than chance level. The authors argued that a possible source of difference between their subjects and the successful guessing of those university students in the other experiments might have been the minimal prior language learning of the "cosmopolitan" undergraduates, i.e., that their linguistic sophistication, compared to the Navajo subjects, might have given them some advantage in guessing. Atzet and Gerard remarked that if phonetic symbolism existed, so far it had been demonstrated only in English. Moreover, onomatopoeia should not be considered as a reliable source of phonetic symbolism since words originally generated by it were also subject to language change. Thus, they concluded, apparent phonosymbolic overlap

---

[29] In the forms: *Which of CD means A? Which of CD means B? Which of DC means A?* and so on.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

between languages would be "probably minimal... and coincidental" (Atzet and Gerard 1965: 529), and that the existence of universal phonetic symbolism was highly questionnable.

In a survey of several word-matching experiments, Taylor and Taylor (1965) made a number of generally interesting obsevations, including criticisms of other researcher's experimental methods and conclusions, and resulting theoretical claims about phonetic symbolism. They made a similar point to Atzet and Gerard's, that where subjects had been exposed, even minimally to the test languages, this might be a factor in correct matching (this point is repeated in Taylor 1965b). Taylor and Taylor argued that if phonetic symbolism were phoneme-based, it could not be universal, since a non-distinctive feature of a phoneme in one language might be distinctive in another.[30] However, in suggesting that the unit of phonosymbolic analysis might be "more gross than the phoneme" (Taylor and Taylor 1965a: 415), they also advanced the idea of phonetic symbolism attaching to certain cross-linguistically distinctive features, such as vowel height or anteriority.

They noted that the use of words in the matching technique was problematic. First, there was the problem of translation bias, i.e., preference for a translation by an item which corresponded markedly in form with the original, rather than providing the best meaning equivalent. This wouldl not be ruled out by the translator's unawarenss of the use of the

---

[30] For example [k^h] is a member of /k/ in English, while /k^h/ is a separate phoneme in Korean.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

experiment: she or he might still choose 'clang'-associate words. Secondly, the phonetic complexity of words meant that the causes of correct matching would never be easy to trace: any number or combination of contrasts might provide the dominant connotation. Also, both auditory and visual methods of stimulus presentation carried drawbacks. The auditory mode allowed for matchings to be influenced via the greater expressiveness of the speaker's voice on certain items. Moreover, in cross-linguistic studies, there was the question of whether pronuncations should be linguistically neutral or native-speaker-like, either of which might have an effect on guessing. Visually-presented matchings might be affected by the word-lengths of items, and by the forms of the graphemes used, whether of the Roman alphabet or other writing systems. Finally, findings of 'correct matching' might be questionnable; since for most studies, correctness scores reported were 50%-60%, these might not indicate guessing any better than chance.[31]

The authors concluded by advancing their *feedback theory* of phonetic symbolism: it was the result of lexical clustering (due originally to chance: a sampling fluctuation in the random distribution of sounds over words), and was therefore language (or language-family)-specific.[32] Speakers would be influenced by a cluster occurring in lexicon, and it, in turn, would be influenced by them, via further word-formation. They argued that every

---

[31] The overall percentage of correctly-matched pairs (the usual measure reported) is a compound of two factors: amount of agreement between subjects and number of pairs matched more often correctly than not. Thus 75% correct matching could be produced a) if subjects always agreed and guessed 75% of the pairs correctly or b) if each pair were correctly guessed by 75% of the subjects, opposite possibilities which would lead to two different interpretations of the data. In lists of twenty to thirty word pairs (the normal length for these studies) "*at least 70% of the pairs must be matched more often correctly than not* for the pair matching result to be significant at the .05 level for a single list" (Taylor and Taylor 1965a: 424, emphasis added).

[32] For a detailed description of lexical clustering, see Bolinger (1965).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sound would not simply become associated with one antonymic pole of every conceptual scale (e.g., *dark/bright*): the scales were not separate, relations between them being culturally determined. Instead, sounds would be associated to a lesser or greater degree with a number of related concepts; cultural patterns would determine strong polarization of certain sounds on certain scales.

Weiss (1966) measured the effects of matching one Japanese word with two English non-antonymic words (sensation-related adjectives, drawn from Roget's *Thesaurus*). Working from a consensus of three translators, twenty-eight Japanese items were selected. Subjects were 290 undergraduate (English-speaking) psychology students, with no knowledge of Japanese or related languages. Significant agreement was reached for 23/28 pairs, but correct matching for only 16/28 pairs (i.e., 57%—less than would be needed, according to Taylor and Taylor (1965), to support Weiss' claim of successful significant mactching).

Weiss observed that he had inconclusive evidence as to whether (English, polysyllabic vs. monosyllabic) word-length was a factor in matching. But he claimed his results suggested that not just antonyms or scales but non-bipolar meaning categories (sense experiences, such as tickling) could become associated with sounds "characteristic of those experiences," (Weiss 1966: 103) and vice versa. Moreover, a single sound might become associated with many different experiences and the connection between sound and meaning-category might then become incorporated into words denoting the category. Universal phonetic symbolism would be generated when across languages, firstly, meaning-categories were shared, and, secondly, connections between sounds and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

meaning-categories were shared. Where in more than one language, non-cognate, non-onomatopoeic words for the same experience were found to have structural similarities, these might constitute evidence for universal phonetic symbolism.

Taylor (1967) responded critically to Weiss's (1966) study, which gave her the opportunity to illustrate arguments agains the theory of universal phonetic symbolism. Thus, she questioned Weiss' apparently arbitrary decision of what constituted 'structural similarity' in words, observing that for any pair, the wrong Japanese word could have been structurally similar to an English word. She also observed that in Weiss' word-lists (Japanese and English) the potentially influencing factors of compounding, suffixation and grammatical class had not been taken into account. Taylor challenged Weiss' assumption that onomatopoeia could not occur language-specifically, and noted that the borderline between onomatopoeic and non-onomatopoeic varied with individuals. She concluded that if the determinants of guessing were reconsidered, then Weiss' results and conclusions were dismissible: the existence of universal phonetic symbolism was not supported, unless confined to imitative words with obvious sound correspondances.

Siegel, Silverman and Markel (1967) focused on the effects of presentation in word-matching experiments. Replicating the original Brown, Black and Horowitz (1955) procedures with Hindi, Czech and Chinese (i.e., matching pairs with pairs), they wished to determine if correct guessing was dependent on visual clues, as suggested by the findings of Brackbill and Little (1957) and Atzet and Gerard (1958). Sets of approximately thirty subjects (with equal numbers of males and females) were tested

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

under two conditions: auditory and audio-visual presentation of stimuli. Significantly correct guessing was found to be higher for every language under the audio-visual condition, although (in contrast to the Brown et al study), Chinese and Czech were also guessed significantly better than chance on auditory-alone presentation. Siegel, Silverman and Markel concluded that orthographic cues had indeed played a part in Brown et al.'s study, and that the lack of antonyms in Brackbill and Little's (1957) experiment might have accounted for differences in results from their own. They also speculated that where presentation was auditory alone (as in Atzet and Gerard's (1958) experiment), phonosymbolic effects might be obtained via the subject's *visualizing* the word; thus, subjects like Atzet and Gerard's that lacked a written language would be deprived of the means of symbolizing, and this would explain their inability to match correctly.[33]

Slobin (1968) set out to answer some questions about the nature of phonetic symbolism in natural languages: whether it was restricted to sensible referent antonyms or to size terms only, and whether there was a difference betweeen judgments along superdimensions of the Semantic Differential. More specifically, he wished to discover if antonyms loaded on the Potency and Activity factors would generate phonetic symbolism, while those loaded on the Evaluative factor would fail to do so. Seven antonym pairs from the three superdimensions were used; all terms were cross-classified for sensible reference (e.g., *happy/sad* were non-sensible), and for size. The matching languages used, Thai, Kanarese and Yoruba, were unrelated to each other and all non Indo-European.

[33] This improbable idea, they felt, would account for Brown, Black and Horowitz' (1955) findings that audio-visual presentation obtained no better scores than visual alone.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Items were back-translated to check accuracy, and speaker expressiveness was controlled for during the recording of items.[34] Subjects were forty-six female undergraduates (*hear and read* procedure), who matched English to foreign pairs using all three languages, and forty-six male and female students (*read only* procedure) who matched English-Thai and English-Kanarese pairs. The *hear and read* subjects were also asked to specifiy whether they had studied foreign languages, and if so, for how long.

Slobin claimed his results to show a percentage of correct guessing which greatly exceeded chance; however, the overall number of signifiantly correctly matched pairs were well below 50% for Thai (both conditions) and Yoruba (*hear and read* only).[35] Significantly correct translations were found for all semantic domains, the most successful being those relating to the Activity superdimension. Phonosymbolic responses were found in non-sensible meaning categories and ones not related to size: thus, it appeared that the phenomenon was not restricted to, e.g., Newman's (1933) domains of brilliance and size. A correlation was found both between foreign language study, and length of study, and successful matching. While generally, matching with written stimuli was successful, disagreements between the two groups over certain Thai pairs suggested that aural symbolism (related to tonal contrasts) might also have played a role. Slobin claimed the tendency to agreement between his subjects gave evidence for American English-related phonetic symbolism, and the cases of correct guessing, for universal

---

[34] Several pairs were later left out of the analysis because they were of the form *X* vs. *not-X* (e.g., the English pair *pleasant/unpleasant*).

[35] These were for Thai *hear and read*, 43% (9/21 pairs), and *read only*, 38% (8/21 pairs), and for Yoruba *hear and read*, 33% (5/15 pairs).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phonetic symbolism. However, he also speculated that the greater success in matching of those experienced in foreign languages might have been due to their experience with other phonosymbolic systems, i.e, that this result gave evidence for Taylor's (1967) feedback theory.

Gebels (1969) reported confusingly on an experiment in which twenty-two antonym pairs of a sensory nature were translated into Old Hebrew, South Malaita, Kiwi, Tongan and Finnish. Subjects were fifty monolingual (Australian) English speakers; they were simultanously (sic) presented with ten words, aurally (sic) and visually as many times as required. Items were both members of an antonym pair in the five foreign languages, which were to be placed on one scale running from 12— to 12+ whose endpoints (presumably) were their English antonymic equivalents. Thus for the English pair E_{a+} and E_{a-}, subjects would be asked to scale OH_{a+} and OH_{a-}; SM_{a+} and SM_{a-}; K_{a+} and K_{a-}; T_{a+} and T_{a-}, and F_{a+} and F_{a-}. Subjects were not told that the items which they were scaling were antonymic pairs or belonged to actual languages.

In obtaining results, for each scaled pair, the left value was subtracted from the right value; thus, it could be determined whether words had been guessed with correct polarization and symmetrically. However, Gebels did not apparently care to specifiy except in the vaguest terms[36] whether significant degrees of matching in the correct direction took place for any or all languages, or what the distribution of correct matchings

---

[36] "The present experiment suggested positive findings" (Gebels 1969: 324).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were vis-à-vis the languages. The only data provided indicate an overall skewing towards evaluating words in a positive direction; unfortunately the conclusions which Gebels draws from this interesting finding are perfectly unintelligible.

Klank, Huang and Johnson (1971) explored the possible linkage between phonetic features and the superdimensions of Potency, Activity and Evaluation drawn from Osgood, Suci and Tannenbaums's (1957) Semantic Differential researches. They argued that most of the earlier studies, consciously or not, had chosen antonyms to be tested which were subdimensions of one of the superdimensions, and the universal evaluability of the superdimensions—and thus of the antonyms chosen—was an essential factor in producing correct matching. Assuming the positive results of earlier studies to indicate phonosymbolic universals, this therefore indicated the presence of phonetic cues which were constant across all languages and which could be tied to the superdimensions.[37]

Six hypotheses were drawn from these premises: 1) That antonym pairs classifiable as belonging to a superdimension (e.g., Potency: *strong* vs. *weak*) would be guesed better than chance; 2) That non-antonym pairs belonging to the same superdimension would be satisfactorily guessed, provided they represented opposite ends (e.g., Potency: *strong* vs. *soft*) because a super-polarity would be operating, with the same phonetic contrast functioning for both scales; 3) That non-antonym pairs belonging to two superdimensions (e.g., Potency: *strong* vs. Evaluative: *good*) would be guessed less well

[37] For example if vowel height triggered positive judgments on one subdimension of Potency (e.g., hardness), it should work similarly for another subdimension (e.g., strength).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

because subjects would have to scan items for two different phonetic cues; 4) That non-antonym pairs belonging to the same superdimension would be guessed less well if they represented the same ends of their subdimensions (e.g., Potency: *strong* vs. *hard*): subjects would be scanning for phonetic contrasts, whereas the same phonetic cue should be triggering the (superdimensional) phonosymbolic judgment; 5) That, assuming universal phonetic symbolism, all of the above should produce the same rates of success, whether the matching was English-foreign or foreign-foreign; 6) That successful guessing for antonym pairs *not* belonging to one of the superdimensions should be more difficult than for conditions 1), 2) and 5).

A first experiment tested eight antonym pairs belonging to the superdimensions (four evaluative, two potency and two activity) and eight which did not. The terms were translated into Japanese, Tahitian, Czech, Hindi and Chinese; where items coincided with Brown et al.'s (1955) materials, the original translations were used. Presentation was visual only, with word-pairs arranged to test each hypothesis; for example, Condition Five translated the Condition Two into foreign-foreign pairings. Subjects were 240 undergraduates, divided into six equal groups. Highly significant results for the Conditions One, Two and Five bore out the hypotheses corresponding to them, as did nonsignificant results for Conditions Four and Six. However, by far the highest proportion of significantly correct guessing was found for the Condition Three (non-antonym pairs from two superdimensions), in direct contradiction of predictions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Attempting to explain this result, Klank, Huang and Johnson observed that if it were due either to factors such as translation bias or word-length, these should have also produced significant results for Condition Four. Could it have therefore been due to some sound-meaning link which allowed correct matchings without the antonymic pairs' having any segments in common? To test this possibility, they asked forty-seven English-speaking students to produce lists of words synonymous or associated with three pairs of antonymic terms, each belonging to one of the three superdimensions. All items produced by two or more subjects were phonemically analyzed, and total production frequency per superdimension was computed for initial consonants/vowels and final consonants. Consonants were also analyzed for features of place and manner, and vowels for those of anteriority and height The ratios of phoneme production per superdimension were then compared with those obtained from 276 Taiwanese students, who created similar lists.

The experimenters were aware that phonemes would not be identical from language to language, and so were examining the correlation of features with the superdimensions. They claimed two important trends. First was a link between initial vowels and Potency (represented by *large/small*), with features of height and frontness denoting smallness, and lowness and backness, largeness. The horizontal dimension was found to be more active in the Chinese data, and the vertical more in the English. Second was a link between initial consonants and Activity (represented by *fast/slow*), slowness being associated with stops and quickness with fricatives. However, since only one antonym pair (i.e., subdimension) was tested per superdimension, it is not clear how generalizable these results really are. No phonetic trends were found relating to Evaluative (represented

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

by *good/bad*); the authors speculated that this might reflect a difference in judgment strategies for this superdimension, based on whole-word Gestalten rather than on the features of individual phonemes. They concluded that their results supported the hypothesis of universal phonetic symbolism.

Kunihara (1971) studied the effect of expressive voice on correct word-matching. Stimuli were twenty-three pairs of Japanese antonyms (in Romanized spellings) matched with one English word. Presentation was under three conditions 1) visual plus monotone aural 2) visual plus expressive-voice aural 3) visual only. Subjects were 195 undergraduate psychology students, with approximately equal groups responding to each presentation type. Significantly better than chance guessing was found for all three types, and Conditions One and Three gave similar results. However, results for Condition Two were notably better, and it was concluded that the expressive quality in the speaker's voice did indeed have a positive effect on correct guessing.

Issues brought up by word-matching experiments. A survey of experiments of this type indicates above all the great complexity of their design: any of a number of variables may have influenced matching, and results obtained thus have a general air of unreliability. Firstly, in examining procedures, we see that translation bias is one problem which was never competely solved (although Brackbill and Little (1957) made a valiant attempt). How to ensure the choice of an L2 word which corresponds to the original only in a meaning-related nonphonetic way? What are the criteria of a context-free accurate translation of a word? Translation bias cannot necessarily be avoided by concealing the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

reason for the translation, or even necessarily by a consensus of translators or by back-translation, since if the phonetic resemblance between two words is strong enough, it may influence translations done for whatever reason, and could affect more than one translator. Moreover, if the translator is told to avoid phonetic resemblances, he or she may be biased in the opposite direction, which is also a problem.

Secondly, other influences could have crept in via stimulus presentation. If the mode of presentation was visual, then word length and/or grapheme shape(s) may have affected judgments, while if the mode was aural, then the presenter's expressiveness of voice may have influenced choices. And behind these issues is another: that of determining how far phonetic symbolism is aural/oral and how far visual. Does it vary according to individual temperament, and how far is the literacy of the speaker a factor? Another question is whether exposure to foreign languages helps in correct guessing—and if so, how?

A third group of issues concern semantic context. How far is it necessary to rely on contrastiveness? Must antonyms be used, and if so, should words be presented in triads or quartets? Or should single words be matched, with judgments confined to sameness or difference, and in that case do these even need to be attributes? Then, there are questions about choice of attributes to be tested. Should these be confined to "sensible referent" qualities such as size and brightness? At what point is an attribute too abstract, or too concrete, to produce correct guessing? Should words be matched only along their real meaning-dimensions, and if not, what are we truly testing? Are many different attributes

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

subsumable under one of Osgood et al.'s (1957) Semantic Differential superdimensions, so that where we think we are testing size, we are in fact evaluating one type of Potency?

Fourthly, there are issues concerning the words to be matched. Should these all be L2-to-L2, or are L1-to-L2 matchings admissible? Most important, what aspect of the stimulus might be triggering agreements in guessing, and how could this be determined? The central problem is that words are complex entities: agreement may be due to lexical associations, such as the occurrence of a phonestheme, like the initial *gl-* of *gleam*, *glide*, *glitter*, or a chance part resemblance to one of the words denoting the meaning-category along which matching is to take place. Alternatively, the trigger might be purely phonetic, but be any one of (or combination of) a number of elements in the word: the initial segment, consonant or vowel, the stressed vowel, the final segment, the arrangement of consonant or vowel types over the whole word and so on. Nor does it follow that phonetic features need to be the same in both words for matching to take place. Having obtained correct guessing, how could the experimenter determine its causes without a costly series of further experiments?

Finally, as was pointed out by Taylor and Taylor (1967), interpretation of results also presents serious problems. Significant but incorrect intersubject agreement indicates the possible effect of, for example, English-specific phonetic symbolism;[38] only significantly

[38] Here, we are assuming English-speaking subjects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

agreed-upon *correct* guessing—and on a majority of the test items[39]—can be held to indicate universal phonetic symbolism. Otherwise, all of the other possible influences need to be ruled out, and we return to the array of difficulties described above.

In sum, there were a daunting number of theoretical questions to be settled and potential influences to be screened out in the word-matching approach; as it was applied, it appears to have led only to very questionnable results and over-grandiose conclusions. In hindsight, it seems that essential problem was the attempt to capture subjective and objective phonetic symbolism in one study. We now have the computerized resources to make very large lexical investigations in which sound-meaning linkages in a language can be established or disproved, via comparison with its normal distribution of sounds. There is no reason why this should not be done cross-linguistically, and for several non-cognate languages. Thus, the largest question raised by these experiments can be settled without the dubious apparatus of word-matching.

Two or three of the studies surveyed (McMurray 1960; Taylor and Taylor 1965; Klank, Huang and Johnson 1971) point the way to a less problematic approach: that of exploring the phonosymbolic effects of feature contrasts. Since we need to examine such contrasts one by one in order to isolate their potential effects, such an approach brings us back to the original insight of Sapir (1927). By using artificial-word stimuli, he set aside the problems of lexical influence; moreover, he focused on a single phonetic object via the

[39] It will be recalled that Taylor and Taylor estimated a significant majority to be as high as 70% for the average-sized word list in this experiment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

minimal pair technique which, as far as possible, neutralized the surrounding phonetic context. Forms of minimal-pair design, simpler and less ambitious than those of the word-matchers, are found in a number of experiments which will be surveyed in the remaining section of this chapter. Stimuli may vary in length from individual segments through syllables and artificial words to sections of discourse (prose and poetry) where one phonetic aspect (vowel or consonant) has been systematically altered and contrasted with another. What links all of the following studies is the attempt to examine specific features as generating phonosymbolic attributions, i.e, to locate subjective phonetic symbolisms without reference to the lexicon.

5.2 Minimal pair experiments.

Hevner's (1936) investigated the influence of four phonetic elements (vowels, consonants, inflection and meter) on spoken poetry. A preliminary study took sections of four real poems that had contrasting emotional effects and deprived them of meaning by changing all their consonants while retaining their vowel, syllable and stress patterns intact. The resulting artificial poems were then evaluated along eight semantic dimensions, each composed of a cluster of eight adjectives. Recordings of all four artificial poems were made and presented to160 students with some experience in this kind of experiment, who were told that the samples were in an unfamiliar language. This initial study did not treat the emotional effects of all four phonetic elements separately but combined them, resulting in global scores. Results showed remarkable agreement between subjects and were consistent with the feeling tone of the real, original poems—

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which argues that the vowels may carry the emotional quality of the poems in major part.[40]

Stimuli for Hevner's main study were also artificial poems, but these did not have literary originals. She designed a neutral basic form which could be manipulated to yield sixteen systematically varied versions, i.e., forms of poetic minimal pairs. Two contrasts were given for each of the four phonetic elements tested; what is of note here was Hevner's contrastive groupings of segments. Vowels were divided into the categories of "light, open, thin, front" (comprising [iy], [ey] and [ay]) and "dark, full, rounded, back-of-the-mouth" (grouping [uw] [ow] with [a], which is not a rounded vowel), and consonants as "smooth" or harsh" (Hevner 1936: 427).[41]

In all, 620 judgements were obtained. Results indicated that meter made the largest consistent emotional impression on subjects, followed by consonant type, and then inflection; least influential was vowel type. However, one advantage of using an eightfold attributional scale was the finding that all of the four phonetic variables, even the least influential, "had some effect in different areas of feeling tone" (Hevner 1936: 430).[42] The

[40] Such findings recall Macdermott's (1946) claim about the expressive primacy of vowels in poetry.

[41] The division of vowels was based largely on the traditional front/back, bright/dark, large/small contrast as noted by many earlier authors (Tolman (1887, 1904), Jespersen (1922, 1933), Sapir (1927, 1929) and so on). Consonants appear to have been assigned somewhat idiosyncratically to the two groups on a general basis of plosiveness versus sonorance; glides were, however, excluded from the "smooth" class, while fricatives varied depending on what final consonant they were used with. The "harsh" sounds comprised voiced and voiceless velar stops, the voiceless alveolar stop, the voiceless palato-alveolar affricate and the sequence [tz] (not an English phoneme—possibly [ts] was meant).

[42] For example, vowels were found to be more effective in "discriminating forcefulness from grace" while meter was much superior to vowels "in determining the happiness or sadness of the poetry" (Hevner 1936: 431).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

difference between the vowel findings in this study and that of the preliminary one, where vowels carried much of the feeling tone, may have resulted from the stimulus design of the main study rather than with what was supposedly being tested. Those emotional effects produced by the vowels were as might be expected; front vowels were "light, leisurely, serene and romantic" while back vowels were more "dignified, emphatic and serious." However, there was one intriguing result; inconsistency was shown in judgements of the "harsh" consonants as "sprightly, merry, jovial and quaint," but also as conveying "heaviness and gloom" (Hevner 1936: 434). This was explained as either as due to the "harsh" group's inclusion of two or more kinds of sounds,or to the multi-featured articulatory/acoustic nature of those sounds as provoking discrepant attributions.[43] The interest of Hevner's work lies in the use of her multi-dimension scale, which allowed some phonetic elements to score high on some dimensions, even if they scored negligibly on others. Moreover, by using poetic minimal pairs she was attempting a systematic investigation of phonosymbolic effects in a linguistic area where they have long been claimed to hold sway.

Research by Eberhardt (1940) investigated the question, first raised by Newman (1933), of whether acoustic or articulatory experience is the more responsible for phonetic symbolism. Eberhardt's subjects were were deaf children who had been taught oral (i.e., non-signed) speech. A design that matched sounds to meanings was used in two

---

[43] See below for a similarly ambivalent evaluation of front vs. back vowels, found in Stevens, De Vito and Isaacson (1969).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

preliminary tests. Children were taught the pronunciations of a number of pairs of new words and asked to say which pronunciation matched one out of a pair of meanings.

In the first test, eight English word pairs were used, five fully antonymic, with six subjects (aged between six and eight). Since some children could master the pronunciation of only a part of the series, out of a possible forty-eight trials less than half were made. Given the smallness of the subject sample and the restricted number of trials, although results were encouraging (17/21 'correct' responses), their significance is anecdotal rather than statistical. But what is of interest is Eberhardt's description of "the matter-of-fact way in which children made their choice. *It was always as though they were using names that had been ordered from the beginning, never as though they were deciding or guessing how to assign them.* One felt.... from their manner of responding... simply that it *was* so" (Eberhardt 1940: 25, emphasis added).

A second pronunciation-to-meaning matching test was done with forty-six children (aged from eleven to eighteen) Overall, responses were not as positive as they had been for the first test; out of the seven pairs tested, only four secured a majority of 'correct' judgements. After analyzing reports made by subjects on the reasons for their decisions, Eberhardt concluded that "choices were made on genuine physiognomic characteristics of the words," (Eberhardt 1940: 27).[44]

---

[44] She suggested some explanations for where subjects "disagreed with the conventional meaning, and therefore possibly with the choices that would have been made by a group of hearing persons." For example, in two pairs, "the position taken by the lips during the pronunciation of the vowel seemed to be the determining factor in the comparison" (Eberhardt 1940: 27) She concluded that "it seems probable that some of the deaf children were influenced relatively more by kinaesthetic than by vibro-tactile sensations

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Eberhardt's main experiment directly compared the results of Newman's (1933) study with results obtained from deaf subjects. The stimulus design of Sapir (1929) and Newman (1933) was used: minimal pairs of artificial one- or two-syllable words, with arbitrary meanings assigned. She first tested the symbolism of size (*large/small*) with fifty-eight advanced students; sixteen weeks later, she tested brilliance (*bright/dark*) using the same stimulus set. Subjects were thirty-two students. Size results indicated a definite patterning of speech sounds in terms of phonetic symbolism with "rough agreement[45] between (Newman's) hearing and deaf subjects" (Eberhardt 1940: 35). On the brilliance comparisons, there was a closer agreement between deaf and nondeaf subjects.

Eberhardt remarked that more specific kinesthetic effects than those suggested by Newman explained the differences between the vowel evaluations of the deaf and the hearing, and that her data "showed cases in which the movements involved in making a sound may carry a symbolic character at variance with that of other sensory data" (Eberhardt 1940: 39). By demonstrating the greater significance of kinesthetic factors in the phonetic symbolism of a particular population, Eberhardt demonstrated that

---

which would more closely correspond to the acoustic sensations of the hearing person... many (judgements) were based on genuine experience of the sensory character of words." However, she also noted some lexically triggered responses "based on superficial verbal resemblances" between the stimulus word and a similar English word. (Eberhardt 1940: 28).

[45] The largest vowel for both the hearing and the deaf groups was [ɔ] and the smallest [ɪ], but by the deaf subjects [uw] was judged as relatively smaller, [ae] was judged close to [ɪ] in smallness, while [iy] was "treated as relatively larger" (Eberhardt 1940: 35) The smallness of [uw], Eberhardt plausibly ascribes to the relative prominence of "lip contraction" for her subjects (Eberhardt 1940: 36), but the explanation for [ae] in terms of a "rigid and tense tongue position" (Eberhardt 1940: 36) is less reasonable. Her results, which place [ae] as smaller than [ay] and [ey], suggest that *length* may be the triggering factor.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phonosymbolisms could vary, depending on modality, i.e., with the suppression of most acoustic triggers apart from vibration. Also, in using deaf subjects, she challenged the conception of speech sounds as inseparable acoustic-articulatory knots, and showed that it was possible to study one of these aspects more or less alone.

Birch and Erickson (1958) tested six vowels, [a], [ɛ], [iy], [ɪ], [ow] and [uw], and three consonants [f], [v] and [n], along Osgood, Suci and Tannenbaum's (1957) three Semantic Differential superdimensions, each represented by a single scale. Potency was represented by *large/small*, Activity by *fast/slow* and Evaluative by *clean/dirty*.[46] Subjects were seventy-four female students (linguistic background unspecified, preseumably American English speakers), and stimuli, fifty-four CVC artificial words (i.e., there were nine consonant-permutations per vowel), orally presented. Findings indicated that [uw] was judged the largest vowel, [a] the slowest and [ɪ] the smallest and fastest. The Evaluative scale did not yield significant results either for vowels or consonants; the authors noted that the high correlation between Activity and Potency appeared to be "largely accountable by a single factor" (Birch and Erickson 1958: 297), but gave no phonetic discussion of why segments were evaluated as they were.

Miron (1961) made a somewhat larger, bilingual (English-Japanese) study also using CVC artificial-word stimuli, orally presented The five vowels tested were [i], [e], [a], [o] and [u] and the consonants, [p], [n], [š], [č]and [g]; compromise pronunciations were

[46] Since there was only one scale per superdimension, it is not clear how generalizable the authors' results were.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

given by the trained phonetician who recorded them. Each vowel appeared with each consonant both in pre- and post-vocalic positions; thus, there were fifty stimuli. Subjects were seventy-nine (American) English-speaking students and forty-one Tokyo dialect Japanese-speaking students and faculty members, all with some English, though with varying amounts. Sixteen scales were given, five corresponding to each superdimension, plus the scale *familiar/unfamiliar* (in order to judge the possible familiarity of each CVC word). Generally, words were judged as unfamiliar, and it was therefore argued that evaluations were not the result of L1 lexical associations. Both groups judged front vowels and consonants as pleasant and weak, and back vowels and consonants as unpleasant and strong. Once again, the Evaluative scales yielded less clear results; for all scales, consonants were judged more consistently than vowels. Miron concluded that his findings gave no direct evidence for phonetic symbolism in natural languages, but that cross-linguistic consistencies in judgments were good indications of universal phonetic symbolism.

Markel and Hamp (1961) confined their study to English. They wished to investigate the independent connotative meanings of five onset CC clusters: [gl], [fl], and [sm], which they felt had strong connotative meanings, and [sp] and [st], which had weaker ones. These were combined in CCV artificial words with two vowels: either [iy] or [a], resulting in ten stimuli. As in Miron's (1961) study, fifteen scales were used from the Semantic Differential, with five scales representing each of the superdimensions, but here the scales were graded along seven points. Subjects were forty-eight students, speakers of American English. Where a lack of significant difference was found for evaluations of an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

onset despite difference in the following vowel, this was taken as evidence of consistent connotative meaning.

Findings showed stronger evaluations for those three CC onsets which the authors had considered to have stronger connotative meanings. They concluded that meanings could indeed attach to smaller-than-morpheme sequences, and proposed the term "psycho-morph" instead of "phonestheme" (Householder 1946) for such units. They felt that this label best indicated the fact that attached meanings were connotative but might not be present with every occurrence. Their conclusions suggested that psychomorphic meanings were based on connotations arising from a lexical cluster (Bolinger 1965). They did not attempt any articulatory- or acoustic-based explanations for their findings.

Taylor and Taylor (1962) performed a set of three experiments, all using CVC artificial words in visual presentation (to eliminate expressive voice bias). In the first, a total of 108 stimuli were tested, combinations of six consonants, [d], [t], [g], [z], [v] and [š] and three vowels, [uw], [ɪ], and [ʌ] (spelled, respectively *oo*, *ii* and *uh* in order to provide uniform word lengths). Eleven monolingual English-speaking subjects (students) evaluated stimuli on a five-point size scale (*tiny/huge*), and definite, consistent rankings were obtained for both consonants and vowels; as might be expected, [uw] was judged as large and [ɪ], small.

Next, the experimenters wished to study the effects of visual word-length on size judgments. They combined four consonants, [g], [m], [t] and [š], and four vowels, [a],

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

[ʌ], [ɪ] and [ɛ] into sixteen CVC syllables, in which each consonant occurred once initially once and once finally with each vowel. To these were added words of the same basic makeup but 'lengthened' by being rewritten in the forms CVhC, CVCC and CVhCC. Subjects were twenty-seven monolingual English-speaking students. Results indicated that syllable length significantly affected the size evaluations.

Finally, a cross-linguistic study was attempted, with speakers of four unrelated languages: English, Korean, Japanese and Tamil. Consonants used were those twelve common to all four: [p], [b], [t], [d], [k], [g], [m], [n], [s], [r], [j] and [š], and vowels were [a], [e], [i], [o], [ʊ] and [u]. These were made into 144 syllables, and given to eighty subjects[47] per language. Four dimensions (warmth, size, movement and pleasantness) were tested separately, i.e., each dimension was given to twenty subjects in each language group. Consistent interlinguistic evaluations were obtained, but the symbolisms were not found to be the same across languages; for example initial *t* and *p* were found to be very small in English and very big in Korean. Also, correlations were found between two scales for one language group, but not for the others (e.g., warmth and movement for the English speakers). The authors concluded that there was no evidence for universal phonetic symbolism, but definite evidence for the language-specific variety. However, a problem not addressed by Taylor and Taylor was the issue of visual vs. auditory phonetic symbolism. Given their methods of presentation, it is not clear whether their subjects

[47] The non-English-speakers were middle-schoolers, and thus somewhat younger students than those speaking English.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were reacting to letter shapes, to internalized sounds evoked by the letters, or to some combination of the two.

In a commentary on the previous year's experiment, Taylor (1963) attacked the word-matching approach[48] and made a case for her 'language habits' or feedback theory. One important point is her stress on the complexity of word stimuli, including artificial words (she remarks that even in CVC forms, any change in the consonantal frame might alter vowel evaluations). A second point is that Anglocentrically-based assumptions of universality have characterized most of the phonetic symbolism studies, including those of Sapir (1927, 1929) and Newman (1933). However, in her own study she had found that the different language groups evaluated sounds/letters differently.

Weiss (1964a) challenged Taylor (1963) and the conclusions of Taylor and Taylor (1962), remarking that Taylor's feedback theory was based on the importance of initial C in the forming of lexical sound-meaning clusters. However, the assumption that this was the only, or most important, triggering position for phonosymbolisms was not, Weiss considered, substantiated. He argued that "structural similarity in historically unrelated words of the same meaning" should be considered as "a potent argument for some process of symbolism" (Weiss 1964a: 456). He also offered the interesting hypothesis that many meanings might be hierarchically associated with a speech sound: this is an insight which suggests the investigation of symbolisms attaching to the different features of a phoneme, although features were not mentioned in the article.

---

[48] For the most developed assault on this approach, see above, Section 5.1 andTaylor and Taylor (1965)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the design of their experiment, Stevens, De Vito and Isaacson (1969) branched out from the construction of CVC's and the reliance on the original scales from the Semantic Differential. Very large minimal pair stimuli were used in both of their two studies. The first set were created via two manipulations of Edward VIII's abdication speech. Systematic vowel and consonant substitutions were made, while respecting the phonotactic rules of English (e.g., no bizarre consonant clusters were allowed). Among the consonants, one feature, not necessarily the same from grouping to grouping, was altered; for example, fricatives and plosives were reversed by manner, but nasals changed by place. Front vowels were backed and vice versa, with height and central vowels remaining unchanged. The two speeches were then played to twenty-five subjects (linguistically unsophisticated students) who were asked to evaluate the speaker's competence, trustworthiness and dynamism, each on four scales.[49] It was found that the back vowel speech scored significantly higher on competence and dynamism.

A second set of stimuli were created, focusing only on vowel features. The experimenters took a basic course-introductory speech and created four versions of it by systematically switching vowels, reversing front/back and tense/lax features. These were played to an audience of twenty-one linguistically unsophisticated student subjects, who were then questioned about the speaker's character (as in the first study), and in addition, the style of the speech and its purpose. In terms of speaker competence, no difference was noted

[49] Corresponding fairly closely to the original Semantic Differential superdimensions, these were superdimensions of a set of speaker credibility measures, designed by Berlo and Lemert (for reference, see Stevens, De Vito and Isaacson 1969: 189).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

between the front and back vowels, but front vowel speaker was considered more trustworthy and the back vowel speaker more dynamic. The lax vowel speaker was considered more trustworthy (expert and just), and there were indications (though these did not reach significance level) that the tense vowel speaker was more dynamic. In terms of the purpose of the speech, there were non-significant indications that lax vowels conveyed positiveness (honesty, wisdom). In terms of speech style, the front vowels were considered good, but the back vowels, emotional, vigorous and personal. Tense vowels conveyed strength and vigor, while lax vowels communicated subtlety, seriousness and earnestness.

The authors maintained that neither theories of phonetic placement nor feedback would account for their findings; however, it can be argued that language-specific social associations to vowel articulations could have played a part. In certain British dialects (possibly also American), the fronting of vowels is found as an upper-class feature of pronunciation, while velarization, involving backing, is a lower-class feature. Similarly, in both British and American dialects, emphasis on the offglide of tense vowels can carry very strong working-class connotations, as can British vowel 'clipping' i.e., shortening, a flavor of upper-class speech. Moreover, vowel pitch and duration (which roughly correlate with the front/back and tense/lax distinctions) have been shown (see below, Tarte 1982b) to be phonosymbolic triggers. It may be that Stevens et al.'s intriguing discoveries are easily reconciled with phonosymbolisms relating to the original superdimensions, i.e., that listener judgments can be predictably manipulated via increased distribution of certain vowel features in discourse. If so, such findings could be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

chillingly applicable to the study of speech-writing, and of the use of voice in public relations and advertizing.

Crockett (1970) made a study using two hundred native Russian speakers, of whom one hundred and thirty were students (aged between seventeen and twenty), and seventy were drawn from all occupations (aged between sixteen and sixty-one). Stimuli were artificial words, mostly of CVC form with a few VC instances (numbers unspecified). Six vowels were used: [i], [e], [a], [o], [u] and [ɨ], while consonants were dental stops and fricatives, and velar and labial stops, and followed normal Russian consonant voicing rules (initial voicing and final unvoicing). Subjects answered questionnaires of two types, but with either type were instructed to base their answers on the sound of the words. Type one questionnaires asked the respondents to match paired artificial words (differing in both consonants and vowels) with near-antonymic pairs of noun phrases (e.g., *little drop* vs. *gigantic tree*) along a five point scale. The middle position was to be used when neither or both noun phrases seemed appropriate. Type two questionnaires gave five pairs of artificial words and asked for the best match with a noun phrase pair; the use of the five-point scale to indicate relevance between items was retained. As there were no significant differences between responses to the questionnaire types, or between student and non-student subjects, these were all pooled in the discussion of results.

Crockett (1971: 109) noted subjects in general had "no difficulty" with evaluations, and even "amazing certainty" in many cases (recalling Hevner's (1936) experience with her deaf subjects). In 70% of cases, the middle position on the scales was not used, i.e., the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

subjects formed a definite opinion about the words. Results indicated that in general, it was the vowels which acted as phonosymbolic triggers: diffuse vowels were judged as small, while compact vowels were large; acute vowels were bright and happy, and grave vowels were dark and unhappy. Thus, vowel height tended to correspond with size, and anteriority with judgments of brilliance and mood in the expected ways (see Tolman 1887, 1904, and so on). Crockett concluded that her results might suggest a lexical basis for phonosymbolisms. However, where parallell symbolisms might exist in a number of related languages, even though each one had independently undergone phonological and morphological changes, an argument could be made that some more general force was responsible for the processes of retention, elimination and word-formation.

The last experiments treated in this survey are all by Tarte (Tarte and Barritt 1971; Tarte 1974; Tarte and O'Boyle 1982a; Tarte 1982b); stimuli were spoken vowels or CVC syllables and/or pure tones. Tarte and Barritt (1971) made a study comprised of three experiments (an initial, and two replications) requiring subjects to label phonetically two types of geometric figures, in a forced-choice test of paired comparisons. Subjects were eight monolingual American English speakers. The phonetic labels were nine CVC syllables; the consonants were, initially, [w], [d], [k] and finally, [s], and vowels were [a], [uw] and [iy]. Presented aurally, each syllable was to be matched with the more appropriate one of a pair of figures in visual presentation (a slide), the same slide being retained through all nine trials. There was a total of eight pairs of figures (i.e., eight slides), varying in four sizes and two shapes (triangular vs. elliptical). Thus, subjects were given 252 trials in all, with randomized ordering of auditory and visual stimuli and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

randomized ordering of all 28 pairings. Right/left subject bias was controlled for by reversing 50% of orderings of the figures on individual slides.

The authors observed that vowels were judged with significantly more consistency than consonants. Preference for vowel matching with the triangular forms ran greatest to least [iy]—[a]—[uw], and with ellipsoid forms [uw]—[a]—[iy], i.e., [a] was considered neutral/indeterminate with regards to shape. However, preferred size matches indicated [uw] as neutral/indeterminate, ranking the vowels as greatest to least in the order [a]—[uw]—[iy]. The authors concluded modestly that their results indicated a "yet undetermined factor" (Tarte and Barritt 1971: 168) allowing AE speakers to agree on size and shape judgments. However, they disclaimed any relevance in their findings to the issue of universal phonetic symbolism.

Tarte (1974) followed his study with one of identical design, but using ten monolingual Czech speakers as subjects. His findings indicated similar responses to those found with the English speakers, including the lack of consistent judgments for consonants. One interesting difference, however, was in the attributions of shape, indicating that [a] and [uw] were treated as similar; thus, rather than the equidistant rankings of the three vowels found in the previous study, preferred matchings for the triangular forms were (greatest to least) closer to [i]—[a/uw], and for ellipsoid forms [uw/a]—[i]. Tarte offered no explanation for this, but it suggests that differences between the phonemic patternings of Czech and American English, namely some sharing of [u]'s and [a]'s features in Czech (e.g., the rounding and/or backing of [a]) may have been responsible.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Ten more Czech subjects were then tested with the same syllable-to-shape matching task, but the range of sizes was reduced to two, resulting in four, rather than eight, figures, and thus fifty-four trials (rather than two hundred and fifty-two). Identical results with those of previous subjects' were obtained, with less subject fatigue. A final test was then run, using a further ten Czech speakers and the reduced number of figures; the original nine [CVs] stimuli were given instead in the form [CVg], in order to reduce possible similarities between the test syllables and Czech words. This produced results similar to those of the previous tests, but it was noted that the vowel-by-shape effect was smaller, while the vowel-by-size effect was larger. Thus, we have a small but clear piece of evidence in support of Taylor's (1963) assertion that alteration in the consonantal frame of a set of minimal pair stimuli can affect vowel judgments. Tarte concluded, once again modestly, that his results might indicate some support for the universality of phonetic symbolism, but that for a solid claim, studies using the same procedure with linguistically unrelated languages were needed.

In a departure from the previous design, but still keeping the focus minimal,Tarte and O'Boyle (1982a) tested the effects of pitch and length on attributions to vowels. Recordings of the vowels [a], [iy] and [uw], uttered in monotone, were made by a (male) phonetician. Four additional types of stimuli were then created by acoustically altering either the frequency or the speed of recordings while holding the other variable constant. The five stimulus conditions thus produced were 1) *fast* (normal frequency/fast speed), 2) *slow* (normal frequency/slow speed), 3) *normal* (normal frequency/normal speed), 4) *high*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(high frequency/normal speed), and 5) *low* (low frequency/normal speed). For evaluation, fifteen bipolar scales from the Semantic Differential were used with seven-point ratings; the superdimensions were represented each by three scales, and the remaining six were measures that might intuitively be used to evaluate speech sounds or voice quality (e.g., *masculine/feminine*, *harmonious/dissonant*) or that related to Tarte's previous studies using figure shapes (*angular/rounded*). *Fast/slow* and *high/low* were among the scales used, and it was felt that evaluations along these dimensions might serve as a check on the effectiveness of the frequency and duration manipulations. Subjects were fifteen monolingual speakers of American English (eight females and seven males: no gender-based difference was found in responses).

From findings on the *slow/fast* and *high/low* scales, it was noted that subjects seemed to have slightly confused the slow and low, and high and fast stimulus conditions, even though the variables had been adjusted separately. Results indicated that across vowels, normal, low and slow conditions were judged generally the same, and that fast and high conditions were judged the same, while the low and fast conditions evoked the greatest contrast in judgment values on all but two of the scales. Thus the low condition scored highest on the Potency scales, and the fast condition the lowest; in contrast, the fast condition scored highest on the Activity and Evaluative scales, while the low condition scored the lowest. Comparing the vowels, the authors noted that [uw] was judged exactly like the low and slow conditions, [iy] like high and fast ones, and [a] as intermediate. They ascribed this to the differing placements of locii of main concentration of acoustic energy in the three vowels. They concluded that frequency was of critical importance in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

carrying phonosymbolic meanings and that duration, though of less interest, also contributed significantly.

Tarte (1982b) focused on frequency jugments in a final study; this compared subjects' evaluations of nine perceptually equidistant pure tones of a constant loudness (ranging from 144 to 8000 Hz.) with evaluations of spoken CVC syllables. These combined the frames [dVs], [kVs] and [wVs] with the vowels [a], [iy] and [uw], and were recorded by a male phonetician, who attempted to equalize intonation, duration and amplitude as far as possible. The same fifteen seven-point Semantic Differential scales were used as in the previous experiment. Tarte's hypothesis was that syllables with [i] nucleii would be evaluated similarly to the high tones, and those with [uw] nucleii, to the low tones. Subjects were fifty native speakers of American English, introductory psychology students; twenty-five judged the syllables and twenty-five, the tones.

While it was found that the consonants were judged significantly different from one another, the vowels received more extreme ratings than consonants, i.e., vowels showed greater evaluation differences between them than did consonants. In terms of the superdimensions, [a] scored highest on the Potency scales, and [iy] lowest, with [uw] indeterminate. On the Activity scales, [iy] scored highest, with [a] evaluated quite close to it and [uw] scoring lowest. On the Evaluative scales, [iy] again scored highest, but here [a] was evaluated lowest and [uw] indeterminate.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The four highest tones were judged similarly to each other, and all were rated differently from the four lowest tones, which were also judged similarly to each other. Subjects were most consistent in evaluations on the Activity and Potency scales, low tones being rated potent but non-active, and high tones the reverse, while on the Evaluative scales, all tones were generally considered 'bad.' Thus [iy] and the high tones were judged similarly for Potency and Activity, but differently on the Evaluative measures; [uw] and the low tones were judged similarly on the Activity scales, while [a] and the low tones were judged similarly on Potency. Tarte suggested this unexpected result might be due to the greater loudness of [a]'s formants, given constant F_0 and amplitude, compared to [uw]'s. He concluded that subjects evaluated pure tones and vowels in a way that was largely in agreement with F_2 and F_3 of the vowels.

Issues brought up by minimal pair experiments. As our survey has indicated, these experiments show more variation in their design than those of the word-matching group—and may even include types of stimulus matching (Crockett 1971; Tarte and Barritt 1971; Tarte 1974). We see that the minimal pair approach does not mean giving up the aim of the word-matching experiments, i.e., enquiry into universal phonetic symbolism. The design and results of Tarte's two earlier experiments (Tarte and Barritt 1971; Tarte 1974), especially the cross-linguistic differences discovered in rankings of [a] and [uw], indicate the value of a series of similar investigations across languages, related and unrelated.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

However, a further isssue remains to be settled. We note that with the exception of Crockett (1970) and Taylor and Taylor (1962), the interlinguistic minimal pair experiments used auditory presentation of word or syllable stimuli, as did most of the intralinguistic studies. Focusing as it does on minimal, single-segment changes—or even changes in features—intuitively, this seems the only realistic choice; moreover, the whole complication of influence via grapheme shape is avoided. But then another decision must follow: whether stimuli should be presented in pronunciations which are native-speaker-like or neutrally-articulated compromises. For native-speaking subjects the advantages of native-speaker stimulus pronunciations are obvious. Equally obvious are their disadvantages for subjects who are foreign-language speakers. On the other hand, if neutrally-pronounced stimuli are presented, how should the evaluations be interpreted, especially differences between those of one group and another?

Stevens, De Vito and Isaacson's (1969) study suggests another line of cross-linguistic exploration, into the effects of the abnormally high distribution of one or more phonetic features over connected discourse. Either their original discourse stimuli could be used with different groups of monolingual, non-cognate language speakers, or equivalent stimuli could be created in non-cognate languages and played to mononlingual AE speakers. The only drawback to this type of project is that if definite emotional effects *were* observed to correspond with phonetic features, results could turn out to be highly applicable to materials in public relations and advertizing. Thus, sellers of products and politics might be offered an even greater capacity for subliminally manipulating ordinary people's attitudes.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Another issue still to be resolved is the comparative strength of symbolic evaluability of consonants vs. vowels. Hevner (1936), and Miron (1961) found that consonants received stronger evaluations, but the findings of Stevens, De Vito and Isaacson (1969), Crockett (1970) and the series of Tarte experiments (Tarte and Barritt 1971; Tarte 1974; Tarte and O'Boyle 1982a; Tarte 1982b) indicated that vowels evoked stronger, more consistent responses. An initial line of further investigation might be with stimulus sets systematically varying both consonants and vowels in CV syllables (as Tarte did, but without adding the post-vocalic position to complicate evaluations further): C_1V_1 compared with C_1V_2 and C_2V_1 and so on. In such investigation, either consonantal features of manner or place would be held constant (unlike Tarte's consonants, which showed variation in place, manner and voicing).

Yet another question relating to studies of vowels is which features to evaluate; for example, the feature of peripherality vs. centrality is treated only in Taylor and Taylor (1962), but is clearly of interest in terms of whether central vowels would evoke extreme or mid-range evaluations. We note that in general, the whole issue of features is extremely tricky. For example, where features are classified by degree (e.g., the five-degree system of English front and back vowel heights) rather than as binary elements, then studying vowel height will very much increase the number of stimuli needed. However, the benefit of such a large stimulus group would be the ability to test front/back contrasts via five separate pairs. But if English vowel height is investigated, the feature of lax/tense (or nondiphthongized/diphthongized) is then also included; thus, the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

complication of final consonants in stimuli creeps back again. If present, they constitute another variable to be taken into account in the evaluations, but if absent, the lax vowel stimuli will violate English phonotactic constraints: this too may influence native speakers' evaluations. We note that Tarte sidestepped these problems—and forwent the benefits—by including only the point vowels in his experiments.

Finally, on the question of evaluating instrument, the Semantic Differential test (Osgood, Suci and Tannenbaum 1957) appears to be highly suitable, and may offer useful generalizations via the superdimensions. Birch and Erickson (1958) reported a confusingly high correlation between evaluations on the superdimensions of Potency and Activity—but in their experiment, each of the three superdimensions was represented only by a single scale. Clearly, the lesson here is not to overgeneralize from one set of subdimensional results. In contrast, Tarte's experiments (Tarte and O'Boyle 1982a; Tarte 1982b) indicated that all three superdimensions (each represented by three subdimensional scales) had psychological relevance, as shown by the agreement between the results of each triad of subdimensions, and by the three differently-ranked evaluations of the point vowels produced by each superdimension. Using different but related measures, Stevens, De Vito and Isaacson (1969) also obtained reliable superdimension evaluations; in their case, four subdimensional scales were used. But there is one disadvantage to evaluating on multiple semantic scales: the risk of fatiguing subjects. It may be worth forgoing ability to score reliable results on the superdimensions for the benefit in making the experimental task easier.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 6

EXPERIMENTAL AIMS, METHOD AND INTERPRETATIONAL APPROACH

6.1 Experimental aims

The primary aim of this study was to establish evidence for or against the existence of phonetic symbolism, determined by examining the responses of a substantial number of subjects when asked for semantic judgements on phones (a group of vowels). Judgments were to be elicited via the use of semantic dimensions: scales of the type used in Osgood, Suci and Tannenbaum's (1957) Semantic Differential Test. If certain qualities were attributed to the vowels with greater than chance agreement, this would be taken as indicating that a consistent phonosymbolic response was at work. A related aim was to discover whether the subject's native language would have an effect on judgements, i.e., whether any phonosymbolic response discovered was language-specific or universal. Accordingly, four language groups were selected (speakers of English, Korean, Japanese and Spanish); 100 subjects from each group participated in the study. The secondary aim was to determine which phonetic or phonological features (taking "feature" in its broadest sense to mean 'noticeable characteristic of a speech sound') might have triggered any consistent phonosymbolic response. Triggering features could then be compared with those (such as anteriority) considered to have been the basis of subjects' responses in the historic studies of Sapir (1929) and Newman (1933).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Using scales of the Semantic Differential type gave rise to a number of subsidiary questions. One was about dimensional clustering: would qualities such as brilliance, energy and size cluster in a consistent way, so that judgements on their related scales could therefore be said to predict others? For example, according to Tolman (1887, 1904) dark vowels are large and bright vowels small. Could a vowel be judged as large *and* bright, and might this be a language-specific issue? A second question was about variation in consistency of response as it related to the semantic dimensions. Might any one scale evoke more substantial agreement than another? For example, would there be the same degree of consensus about about a vowel's brilliance as about its size? A third question was about judgments evoked by the use of antonymic semantic scales, i.e., the splitting of a dimension (such as brilliance) into its two poles, bright and dark. What effect might be observed by asking for the (apparently) same sets of vowel judgements twice? If subjects were asked how dark a vowel was, would the same evaluations be made as when asked how bright it was—or could the response patterns vary? Would results suggest that judgements were made on a unified scale of brilliance, or rather on two separate scales: one of darkness and one of brightness?

6.2 Experimental method

Subjects. Speakers of American English (hereafter abbreviated *AE*), Korean, Japanese and Spanish participated in the experiment. There were 100 subjects in each of these four groups, ranging in age from 18 to 60. Some 85 of the 100 AE speakers were recruited from

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

undergraduate introductory linguistics classes in the City University of New York.[1] The Korean, Japanese and Spanish-speaking subjects (hereafter referred to as the *ESL* subjects) were drawn almost entirely from undergraduate classes in English as a Second Language, at various branches of the City University of New York, at New York University and at a private English school for Japanese speakers.[2] Among these ESL subjects, those identifying themselves as having native-like English accents were excluded.

Materials and design. Since the goal was to elicit judgments from a large number of subjects, and to economize on time by obtaining data from as many as possible simultaneously, the written questionnaire method was chosen as the quickest and least cumbersome. The questionnaire was designed to be short—answerable in around 20 minutes at most—to minimize subject fatigue. Since the aim was to elicit attributions to speech sounds, and the hypothesis was that such judgments come from a substrate of phonetic-attributional associations formed very early in linguistic development,[3] it would have been counterproductive to make participants feel that they were laboring away at a tiresome task. On the contrary, following Mandelker (1982), who told her subjects that they were evaluating a new art-form, the goal was to encourage a playful, childlike state of mind. Therefore, the questionnaire featured a small number both of semantic dimensions (three), and of vowel stimuli (four). The questionnaire was offered only in English, since all of the ESL subjects, as students of English, were expected to have a minimal English

[1] However, the AE group also included several graduate students in linguistics (all from the City University of New York), a professor of linguistics, and three personal friends of the experimenter.

[2] Exceptions were one Spanish-speaking and one or two Japanese and Korean acquaintances: these were a Venezuelan architect, Korean and Japanese graduate students in linguistics, and one Korean and one Japanese artist.

[3] For an illuminating discussion of this point of view, see Fónagy (1979, 1983), discussed in Appendix 6.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

proficiency. This avoided translation problems with the adjectival terms used on the scales; it also prepared subjects for evaluating AE vowels.

The dimensions examined were brilliance, energy and size, which were implemented as three pairs of antonymic scales (respectively, bright/dark, active/passive and large/small). The dimensions of brilliance and size have been traditionally used in phonetic symbolism studies from Sapir (1929) and Newman (1933) onwards; the primary reason for including them were so that results of this experiment could be compared with findings obtained by these pioneer researchers.

In addition, since scales of the Semantic Differential type were being used, it was considered desirable to include scale-pairs representing Activity and Potency, i.e., two out of the three metadimensions discovered by Osgood, Suci and Tannenbaum (1957). Thus, besides their historical connections, brilliance and energy were also included as exemplars of Potency, while Activity was represented by the energy scales of active/passive.[4] Given their similar factor-loadings, it was expected that bright/dark might evoke a response pattern similar to that of large/small and different from that of active/passive. No pair of scales was chosen to represent Evaluation (Osgood et al.'s third metadimension), because, following Solomon,[5] it was considered that judgements along this dimension would be "usually of an idiosyncratic nature"

---

[4] Brilliance and size were found to be highly factor-loaded for Potency, while activity was a "pivot" scale with maximum factor-loading on the Activity metadimension of the semantic differential (see Osgood et al., 1957: 62-63).

[5] In his study of attributions to passive sonar sounds, Solomon (1959a) found that five semantic differential scales highly loaded for Evaluation did not produce consensus in semantic judgements (see Chapter 4, Section 4X).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tending only to reflect "the unique experiential background of the perceiver" (Solomon 1959a: 495).

Stimuli. The phonetic objects to be studied were kept as small as possible, so that extraneous influences on them could be excluded or reduced to a minimum. Isolated speech sounds were selected as stimuli; consonantal effects on vowels or vocalic effects on consonants could thus be avoided, and lexical influence on phonosymbolisms would be minimized. Speech sounds of a single major class were decided on, and vowels were chosen rather than consonants, firstly because there were some grounds (literary rather than linguistic) for believing that vowel sounds might be phonosymbolically primary; studies of phonetic symbolism in poetry (Hevner, 1937, Macdermott, 1940, Mandelker, 1982) have focused on vocalic quality as the most important vehicle of emotional tone. Secondly, vowels are easier to produce in isolation.[6]

American English target vowels were chosen for the stimuli. The peripheral vowels [a], [i] and [u] (hereafter referred to as the *point* vowels) were selected as being most articulatorily and acoustically extreme,[7] and therefore most likely to provoke polarized judgements. To contrast with these three extremes, it was desirable to include also an articulatorily "neutral" (mid central) vowel. Although the most centrally positioned, the schwa vowel was excluded as likely to focus subjects on features not conceived of as part of the study, namely those of stress and duration. Instead, [ʌ] was chosen; in descriptions of American English, it is often referred to as stressed

[6] This is due to the fact that (with the possible exception of the voiceless fricatives) most consonants will be articulated with some vocalic offglide, at the very least a small schwalike vocalization; thus, where an isolated consonant is required, a CV syllable is likely to be produced instead.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

schwa (Ladefoged, 1982). However, [ʌ] is by no means always a central vowel for every English speaker. Mackay (1987), for example, classifies it as the mid-low back unrounded equivalent of [ɔ]), and the variability of [ʌ] proved to be one of a number of difficulties in interpreting judgments about this stimulus.

A crucial feature of the questionnaire was that each subject would be asked to make his or her own vowel stimuli; thus, no exterior models in the form of recorded stimuli were employed.[8] This yielded a triple advantage. Firstly, using the subjects' own productions (rather than standardized stimuli) would allow them to access items from their individual phonemic repertoires directly, whereas exterior models which could have interfered with their own inner representations of the vowels. Secondly, it allowed the proprioceptive aspects of the vowels to be part of the evaluation process, rather than their acoustical qualities alone. Finally, on a practical level, self-produced stimuli avoided the need for recording playback equipment and all of the possible complications associated with the use of recordings.

In designing the questionnaire, finding unambiguous written representations for the vowels was the first concern. Clearly, these needed to be in standard English orthography, since the learning of any unfamiliar signals would add to the experimental task, and thus risked fatiguing some participants. But all spelling solutions necessarily have their drawbacks.[9] In English orthography,

---

[7] The point vowels are articulated at the extremes of the vowel space; consequently, their F_1 and F_2 formant values will be the highest (or lowest) of those for all of the speaker's vowels.

[8] For this original suggestion I am much indebted to Dr. John D. Roy.

[9] Any visual sign chosen to represent a speech sound (phonetic symbol, letter or combination of letters, drawn figure) will have its semantic disadvantages, since it is likely to provoke some attributional associations in some subjects. The issue is therefore that of choosing the least association-provoking signs, and it is crucial to pilot the candidate

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

there are fairly consistent representations of the high vowels [iy] and [uw], respectively as *ee* and *oo*, and these were used in the questionnaire. However, the low nonanterior vowel [a] found in AE *hot*, *arm* or *calm* is quite rarely represented by *a* without a following nucleus-associated character (mostly *r*, occasionally mute *l*, as in the above examples). Despite this, *a* appeared to be the best choice to signal [a], since the character *o* was already being used in the digraph *oo* for representing [u]. However, orthographic representations of the vowels needed to be all of the same length, or subjects might well have conceived of the vowel spelled with fewer characters as being shorter in duration, which in turn could have influenced phonosemic judgments. Therefore *ah* was used as the best representation of [a]; for English readers, it is the classic way to write this sound. But this was a compromise, since the postvocalic *h* might run the risk of confusing some ESL subjects.

More problems were encountered in the representation of [ʌ]. In standard English orthography, this vowel is generally signaled either by *u* (as in *cup*, *chum*, *bucket* and so on), or with *o*, (as in *come*, *mother*, *Monday*, *love* and *money*). Even if *oo* had not been already appropriated for representing [uw], use of a single *o* as a signal for [ʌ] might still have proved confusing for subjects, since *o* can also signal other vowels.[10] The only reasonable choice for a stimulus signal appeared to be *u*. But in English spelling there is also an orthographic association between *u* and [uw] (*rude*, *tune*, and so on). Thus, although *u* was the spelling finally employed, this too was a compromise, and might also prove a source of subject confusion. As with the representation of [a], a digraph was needed to equalize stimulus lengths visually, and so *uh* was selected to signal

---

signs well before testing.

[10] Orthographic o can signal [ow] (as in *so*, *go*, *lonely*, *vote* and so on) or [uw] (as in *move*, *lose*, *to*, and *do*).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

[ʌ]. Like *ah*, it is the classic written representation of this sound for English readers, but again, the postvocalic [h] ran the risk of confusing some ESL subjects.

Even though the original decision had been to exclude consonants from the study, on closer consideration, it was found necessary for each stimulus vowel to be preceded by an initial [h]. This was to make the pronunciation task easier for three out of the four groups of subjects, since isolated, stressed vowel-initial syllables are phonetically problematic in English,[11] Korean[12] and Japanese (although not in Spanish).[13] Yet any initial consonant chosen had to be one which would not introduce any extra oral features; i.e., it should modify the following vowel as little as possible, or risk introducing additional potentially phonosymbolic material. The obvious choice was [h], an orally featureless consonant which also presented no problems in its orthographic representation.[14] The only apparent risk[15] was of some portion of voicelessness at the beginning of the pronunciations of the vowels, but since initial [h] was to be constant for all four stimuli,

---

[11] CVC is the (phonetic) rule for English stressed monosyllables. For isolated, orthographically vowel-initial syllables (such as *I, out, aim, owe, eat*), phonetically, the first consonant is a glottal stop.

[12] See Lee (1955: 8) for a description of an orthographic, possibly phonetic, consonantal epenthesis which appears very similar to English glottal stop insertion in the same phonetic context.

[13] For example *u-na*: 'a, one' (feminine) , *e-ra*: 'was' (3rd person singular), *i-ba*: 'went' (3rd person singular), *a-jo*: 'garlic', *o-tra*: 'other' (feminine). The initial vowel is not preceded by a glottal stop, as it would be in English.

[14] The only other orally featureless possibility was the glottal stop, of which there exists no representation in English orthography.

[15] In fact, it was subsequently realized that another element had entered with [h], which had not been allowed for. This was a consequence of the fact that in Japanese, [h] assimilates to following high vowels, thus producing labialized and palatalized variants; i.e., [ç] precedes [i] and [Φ] precedes [ɯ] (Han 1961: 27). Consequently, the requirements for the Japanese subjects to produce [hi] and [hu] for the high vowel stimuli, was in violation of their own phonology and may well have had an influence on evaluations (see Chapter 9, Section...). I am indebted to Dr. Edward Bendix for pointing this out (in a personal communication). I am also informed (by Dr. Bruce Ballard, in a personal communication) that Korean [h] may be similarly affected by assimilation; though the pre-[i] palatal variant is rarer, the pre-[u] labial [Φ] is quite normal.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

any initial devoicing would be consistent over all of the vowels and therefore should not skew the evaluations.

Thus, the final choices of target vowel representations on the questionnaire were as follows: [a] = *hah*; [i] = *hee*; [u] = *hoo*; [ʌ] = *huh*.. In all subsequent descriptions of this experiment, where "stimulus vowels" are referred to, the reader should understand that these were embedded in [hV] syllables. Where the focus is specifically on stimulus spellings as found in the questionnaire, they will be represented as *hah*, *hee*, *hoo* and *huh*; in every other case, the stimuli and/or the target vowels will be represented by the angle-bracketed vowel symbols <a>, <i>, <u> and <ʌ>. This is in order to distinguish the (first-language) vowels of the subjects, which will continue to be represented by vowel symbols in square brackets, from the stimuli and/or target vowels.

Word-lists. In order that subjects would understand which four vowels were to be evaluated without their pronunciations being externally modeled, four 12-item lists of English words were compiled to be administered before the questionnaire (see Appendix 1). Each list exemplified one vowel; for example, the list for <u> was as follows: *shoes, Lewis, grooming, undo, glue, movement, soup, truly, Susan, rooting, bamboo, sooner.* Words as semantically varied as possible were selected, so that no one meaning could be associated with a vowel sound, and thus there would be no meaning-carryover from list to vowel pronunciation. However, as can be seen from the example above, the words on the word-lists were frequently inconsistent with the stimulus spellings on the questionnaire, [16] even though it would have been possible to produce

[16] This was the result of not conceiving of the problem in terms of what the easiest word-lists would have been for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

consistently-spelled lists for three out of the four vowels.[17] Difficulties in generalizing from the word-lists to the target vowels were to add substantially to the preliminary instruction time for many of the ESL subjects, especially for the list exemplifying <ʌ>. It may well have decreased their openness to playing phonosymbolically—in Fonagy's (1978, 1983) term, 'regressing'—by aggravating their nervousness and fear of failing at the experimental task. However, by the time that it became clear that an extra source of confusion had been introduced (after observing many subjects struggle with pronunciation of the word-lists), it was already too late to change the lists without introducing a new design feature into an experiment well under way.

Moreover, since from the outset the design had included ESL subjects yet was entirely couched in and focused on English, the requirement of self-produced stimuli was to carry an inherent contradiction. The projected direct access to the subjects' own phonology would only be perfectly achieved by the AE participants, since the vowels produced by the ESL groups would necessarily be some kind of compromise between American models and their own native Korean, Japanese or Spanish sounds. This *intermediate* nature of the ESL subjects' target vowels was later to raise difficulties in the interpretation of the ESL data; it underlined the cost of attempting to standardize the vowels for different linguistic subject populations without using an exterior model.[18]

ESL students to associate with the target vowels. Rather than giving the maximum range of target vowel spellings, as was done in the experiment, the easiest would have been lists of common monosyllables with one consistent spelling for each vowel—that found in the spelling of the stimuli on the questionnaire.

[17] The exception is [a] in American English. It appears that in this dialect there exist very few words like *father*, where [a] is represented by orthographic *a* alone, without a postvocalic *r* or mute *l*.

[18] Of course, had an exterior model (i.e., recordings of the vowels) been used, the same issue of exactly which vowels to use would have come up in another form—plus the extra decision of using a male or female speaker. Should American English vowels still have been presented? Or should cardinal vowels have been used, at least for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Questions. The questionnaire was to be answered anonymously.[19] Part I was comprised of questions on (1) the subject's geographic/national background, (2) age on first arrival in the U.S., (3) sex, (4) current age, (5) first language, (6) other languages spoken and (7) years spent studying English. However, in the course of the experiment, it became clear that three of these questions, relating to language background, were inadequate. They had not been worded precisely enough to obtain reliable data on the amount of exposure to foreign languges and to English.[20]

Part II of the questionnaire was comprised of those test items requiring evaluations of stimuli. Each featured a stimulus pair rather than a stimulus in isolation, and took the following form:

> *Make the sounds hah-hoo, hoo-hah, hah-hoo,hoo-hah.*[21]
> *Out of hoo and hah which one sounds darker?*

---

the three point phones, even though these would not have been exactly the same as those of any subject's native language? And if the cardinals had been chosen, should a non-peripheral cardinal vowel still have been included?

[19] For a sample questionnaire, please see Appendix 2.

[20] Some ESL subjects, especially the more prosperous, cosmopolitan ones, had difficulty in answering question 2, on the time of their first arrival in the U.S. What was meant by this—a short holiday, perhaps taken many years ago, or the moment when they arrived as permanent immigrants or as students expecting to spend several years here? Question 7 (number of years spent studying English) yielded much more widespread confusion. Should the subject include his or her practical experience of speaking English in everyday life in this country, or only count from the beginning of formal, academic study, perhaps several years later? And if academic only, what level of study was being referred to: classes taken in high school, or even earlier, although possibly very limited and with minimum practice of pronunciation? The solution was to advize subjects to answer 7 by indicating when they first felt they were truly speaking and learning English, i.e., when English first palpably entered their lives. But it was clear that the vagueness of this instruction meant that their responses could not be used as a stable, quantifiable measure of exposure to the language. Likewise, question 6 was confusing for the English-speaking subjects: there was no way of reporting the intensity or casualness of any foreign language study, or the effect of other languages spoken in the home. Worse still, there was no question included to weed out those bilinguals with native-speakerlike pronuncition of English. Therefore, at every session, it was necessary to ask all participants who considered themselves as this type of bilingual specifically to mark this as part of their answer to question 6.

[21] Instructions were to pronounce the pair of syllables four times in both possible orders in order to neutralize any effect of order of presentation of the stimuli.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The items were constructed in this way in order to make the evaluation task as natural as possible, based on the informal observation that questions of the type *Which of these two syllables sounds more X in quality?* produced notably more confident phonosymbolic responses than *How X in quality does this syllable sound?*[22] Moreover, the investigation had been conceived of as exploring those phonosymbolisms which might be attached to four[23] items drawn from the subject's phonemic inventory, items that are considered to derive their identity through opposition to other items in the system (de Saussure, 1959: 120-122). Therefore, getting the subject to focus on vowel-contrasts[24] was expected to be a useful means of triggering semantic attributions. Finally, a substantial minimum number of judgements per stimulus were required in order to gather a large enough quantity of responses for reliable analysis; pairing each stimulus with the three others had the additional advantage of tripling the number of evaluations to be made of it.

Each exploration of dimension was focused in two ways:

Which of these two sounds is the brighter/larger/more active?
Which of these two sounds is the darker/smaller/more passive?

This was to determine whether if asked the (apparently) same evaluative questions from the opposite end of the dimension, they would generate responses in the same direction and of the

---

[22] The latter question is of course the classic semantic differential task, with evaluations measured on a seven-point polar scale (Osgood, Suci and Tannenbaum, 1957: 25-30).

[23] However, for the Japanese and Spanish speakers with five-vowel systems, it is by no means clear that four psychologically separate items were being contrasted in all of the questions, since AE [ʌ] did not correspond to a phoneme in their own first language. Moreover, authorities are divided as to whether a close equivalent for AE [ʌ] exists in Korean (see Chapter 9, Section...). Thus, it is a matter of speculation as to exactly what was being contrasted for the ESL subjects in questions involving the comparison of <ʌ> with a point vowel stimulus.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

same strength (see above, Section 6.1). Pairing the four stimuli resulted in six combinations (the stimulus pairs <ʌ/u>, < ʌ/a>, < ʌ/i>, <u/a>, <u/i>, <a/i>), and assessing each of the three semantic dimensions on its two component scales yielded six different questions. There were thus 36 vowel judgement questions in total (6 stimulus pairs x 3 dimensions x 2 scales).

The option was also explicitly included for the subject to answer "neither" if the semantic quality in question did not seem relevant to either of the stimuli, or it was otherwise impossible to decide between the two.[25] Certain critics of phonetic symbolism experiments, from Bentley and Varon (1933) onwards, have accused experimenters of 'discovering' symbolisms whose presence was established only by offering subjects forced-choice questions, their arguments being that if subjects had been given the opportunity of making no choice, the claimed symbolism would never have emerged. Thus, it was necessary to give subjects the possibility of opting out if they wished; it could not then be objected that results had been obtained via a kind of evaluational strong-arming.

For the purposes of pseudorandomization of stimulus pairs and scales, the 36 questions were divided into 6 blocks of 6 items. All stimulus pairs were rotated through all scales in every block;

---

[24] This was, paradoxically, using the Saussurean principle of oppositions to question the Saussurean principle of arbitrariness, but the reader will have already observed (see Chapter 2, Section...) how profoundly, from Darwin's Principle of Antithesis onwards, the issue of phonetic symbolism is associated with that of contrasts and polarities.

[25] When examining results, it was found that individual subjects had rarely availed themselves of the *neither* option, perhaps once or twice out of the total of their 36 responses, if that. However, in the case of the AE group, there were eight subjects, who gave very high numbers of *neither* responses, between 7 and 14. Seven of these eight were mature-age linguistics graduate students (aged between 36 and 50), suggesting that their relatively greater age may have been a factor in their responding on a somewhat different basis from the other AE speakers, although their non-*neither* answers were patterned similarly with those of the rest of the group(in fact the AE speakers had the highest overall amount of consensus out of the four). These eight subjects were excluded from the study and replaced because the large number of *neithers* presented a statistical problem for the ranking process.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for example, in Block 1, <a/i> would be presented once, and once in Blocks 2, 3, and so on, with the position in which the pair might appear in any block being random. The same was true for the scales, so that, e.g., one brightness question would be asked in Block 1, one in Blocks 2, 3, and so on. The order of presentation of the dimensions and of the scales in each block was random, but the presentation order of the antonymic scale-pair members of each dimension was not. Thus, if a brightness question were asked of <a/i> in the first block, then the next block would contain its polar opposite (i.e. a darkness question) asked of the same pair.

In the course of the data collection, it was suggested that subjects' understanding of the adjectival antonyms used—*bright/dark, large/small, active/passive*—might be a useful focus of investigation. Accordingly, Part III, comprising two sections, was added to the questionnaire. The first section was only for ESL subjects and asked for translations of each of the adjectives into their first languages. The second section required all subjects to provide associations to the adjectives, with questions phrased as in the following example:

What are three things that you normally think of as bright?

The first section provided important results for interpreting the Japanese and Spanish data, but responses to the second section proved to be too scattered to offer useful information, and this part of the investigation was discarded. Two versions of Part III were offered, with adjectives given in two different orders. [26]

[26] Samples of part III, illustrating both adjectival orders, can be found in Appendix 2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Procedures. Subjects were tested in groups[27] varying in size between 5 and 45, in a convenient, quiet location, e.g., a classroom. With the exception of those groups in the private Japanese language school, the groups were multilingual. The same procedures were always followed.

Administering the word-lists. Each list was taken separately; subjects were asked to read all items on it aloud, focusing on the target vowel sound exemplified. Pronunciations were then elicited of the target vowel, first from individuals and then the whole group. Thus, on the first list, subjects were asked to produce "target vowel number one," which was *hah*. Although the administrator did not model the vowel, she signaled if pronunciation had reached the target or not. As the task tended to make some subjects quite anxious, particularly the ESL ones,[28] care was taken that the instructions were given in a supportive, playful and patient manner. The second word-list was not attempted until it was clear that subjects had made a reliable association between the first target vowel's pronunciation and the orthography used to represent it. The same process was repeated for each of the lists; then several review pronunciations of all four target vowels were elicited in random order from the whole group of subjects.

Administering the questionnaire. When the administrator was satisfied that there was a firm connection in subjects' minds between the experiment's orthographic rendition of the target vowels and their pronunciations, the questionnaire was given out, and subjects were asked to complete the questions in Part I. Before the subjects began on Part II, it was emphasized that this

---

[27] Exceptions were the personal friends of the experimenter and the two Asian-language-speaking artists, who were tested individually, in a domestic environment.

[28] Given the varied spellings on the word-lists, all of the ESL subjects who were beginners in English (or were not truly practiced English readers) were likely to have been experiencing reading difficulties.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

was not an examination and that there were no correct or incorrect answers, responses being a matter of personal taste. Subjects were requested not to refer back to earlier answers, but rather to treat each question as a separate opportunity to evaluate and decide. They were also reminded that although they should make a sincere attempt to respond to every question with an evaluation, if after some consideration they found that neither of the two sounds they were judging seemed to have more of the quality asked about, they had the option of signaling "neither."

Subjects were then asked to read aloud together the stimuli referred to in the first question, to consider the question and to write down their choice. When all the subjects had pronounced the first stimulus pair and indicated that they had written a response, they then proceeded to answer the other 35 evaluation questions at their own individual rates. Finally, Part III (a separate page, stapled to the other three) was completed by those participating later in the experiment.[29] The average time for subjects to complete Parts I and II of the questionnaire was around 10 minutes for the AE subjects and around 10 to 15 minutes for the ESL subjects. Time taken to complete Part III varied more, between 3 and 10 minutes: this appeared to be because the task in the second section, requiring free association, was experienced by some subjects as particularly challenging.

Data types and treatment. Responses to Parts I and II of the questionnaire, the 7 demographic information questions, and the 36 stimulus evaluation questions (the latter hereafter referred to as *test items* in order to differentiate them from the other items on the questionnaire), were entered

---

[29] The reader is reminded that Part III was added only when data collection was already well under way; sets of responses were obtained from 33 English, 50 Korean, 60 Japanese and 26 Spanish speakers.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

into a spreadsheet. Scores were totalled within each subject group for each test item;[30] these totals would form the basis for the two principal types of data: *consensus distribution* and *stimulus rankings*.

Consensus distribution. The first type of data to be extracted from the original scores was the distribution of responses to the each of the test items by each language group. To recapitulate, each test presented a stimulus pair (i.e., one of the following six combinations: <ʌ/u>, <ʌ/a>, <ʌ/i>, <u/a>, <u/i>, <a/i>), and the subject decided which of the two displayed more of the quality specified by the associated scale (i.e., of brightness, darkness, activity, passivity, largeness, smallness). Each subject answering the questionnaire made responses to 36 test items, and on each test was offered three choices. He or she could 'vote' for either one of the two vowels mentioned (an evaluating choice) or 'opt out,' (the *neither* choice). Thus, a language-group's response total for every test item was made up of three separate frequencies whose sum was always 100 (the number of subjects in each group). The frequency of *neither* choices normally varied between 0 and 3; the highest total of *neither* choices for any language group was 6 (found on the Spanish smallness scale). Consequently, the lowest number of evaluating choices evoked by any test item was 94, but the normal range was 97 to100.

The term *statistically significant consensus*, hereafter abbreviated to *consensus*, will be used to indicate that among the responses within a language group to a particular test item, one vowel gained too large a majority of the evaluating choices for this to be ascribed to chance: i.e., the chi

---

[30] See Appendix 3 for a spreadsheet giving the original totals.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

square test returned a significant result ($p < .05$). For each test item, individually, a statistical evaluation took the form of a two-tailed chi-square goodness-of-fit test, against a null hypothesis specifying that the evaluating choices were distributed randomly between the two stimuli; instances in which subjects responded *neither* were discarded. The significance level was set at 5% throughout. Where subjects failed to reach statistically significant consensus on a test (i.e., the chi square returned a non-significant result: $p > .05$), this will be termed a case of *non-consensus*. Trends towards significance (where $.05 < p < .01$) were counted conservatively as instances of non-consensus.[31]

Examination of the distribution of instances of consensus could be approached in several ways. Three variables were involved: language group, stimulus pair and semantic scale. The broadest approach was via the first variable; it was to make a comparison of the extent of proportions of consensus on the 36 test items achieved by each of the four language groups. The extents themselves were submitted to statistical analysis, and it could thus be determined whether consensus occurred often enough to warrant confidence in the claim of a reliable basis for phonosymbolic response. In fact, frequencies for every subject group reached significance in over 60% of all cases.[32] This initial finding oriented subsequent analyses. The comparative rarity of non-consensus meant that both isolated cases and general patterns of their distribution could be scrutinized as giving potentially valuable evidence. Clearly, however, non-consensus data

[31] To give the reader a sense of how large a majority supports a significant chi-square result, consider a test item in which all 100 responses were evaluating choices. A split in responses favoring one stimulus over the other 60/40 occurs with $p < .05$. Splits that are less extreme, e.g., 59/41 through to 50/50, give $p > .05$.

[32] Consensus was exhibited at the following rates: English: 80.5%; Korean: 69.4%; Japanese: 63.8%; Spanish: 75.0%.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(interpreted as indicating the phonosymbolic *collapse* of two stimuli) would provide a different type of information from that provided by consensus.

Apart from language, the consensus/non-consensus distribution could also be approached in terms of the two other variables: semantic scale and stimulus pair. Accordingly, two types of (intralinguistic) consensus tables were set up; these were tables organized by *scale* and tables organized by *stimulus pair*, the latter being especially useful in indicating how pair-based patterns of non-consensus compared across languages. Since the consensus-by-scale tables indicated which scales evoked the greatest amount of consensus and which the least, a number of possible scale-based comparisons could be made, including those between individual scales, dimensions and, in particular, paired antonymic members of a dimension. In the analyses to follow, it will prove useful to employ the term *salient* to refer to the brightness, activity and largeness scales, since in everyday life these are the qualities which generally draw the attention. Their antonymic counterparts, the smallness, darkness and passivity scales, thus become *non-salient*.[33] It was hypothesized that subjects might have greater difficulty in evaluations involving the non-salient adjectives *dark*, *passive* and *small* than they would their salient scale counterparts.

The importance of non-consensus. Consensus on a test gives evidence that one of a pair of stimuli is reliably judged to be larger or darker or more passive, and so on than the other. But what about non-consensus? To take an illustrative example, suppose that a given subject group responds to a given stimulus pair (say, the stimuli <ʌ> and <u>) on a given scale (say, passivity)

[33] For these terms, I am obliged to Dr. Edward Bendix, who suggested them in a personal communication.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with identical 50/50 scores.[34] These scores still do not give direct evidence that the two stimuli were judged equally passive; they indicate only that on this scale, *subjects failed for some reason to differentiate between <ʌ> and <u>*. Yet given the comparative rarity of non-consensus, evidence of this kind of failure to differentiate was in itself considered an important finding. Wherever more than one case per stimulus pair was discovered, reasons for the failure to differentiate were carefully reviewed. In addition, since non-consensus was much more frequent on the non-salient scales, the doubly unusual occurrence of non-consensus on a salient scale was considered to have especial importance, i.e., as carrying extra weight. Interlinguistic similarities and differences in the distribution of non-consensus were especially illuminating, both in clarifying the intralinguistic analyses and also in the discussion of universality (see Chapter 11, Section 11.2).

Stimulus rankings. The second data type extracted from the original frequency totals were stimulus rankings. Application of a metric produced values of a *ranking index*[35] for each of the stimuli, separately for each of the six scales for each language group. The index was useful in establishing, for example, which stimulus was considered by a given group of speakers to be the brightest, and which the second-brightest, the third-brightest and the least bright.

In the questionnaire, each of the four vowels was paired in turn with each of its fellows to produce three test items on a given scale, e.g., stimulus <i> occurred in pairs <i/ʌ>, <i/u>, and

---

[34] Exactly this outcome occurred for Japanese with the pair <ʌ/u> on the passive scale; 'votes' were split 50/50 between the stimuli.

[35] For a table displaying ranking index values, see Appendix 3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

<i/a> in test items relating to brightness. Suppose the vowel in question attracted evaluating choices in its three competitions for that scale (i.e., was judged, say, the brighter of the pair) with frequencies f_1, f_2, and f_3, respectively, while its paired alternative attracted evaluating choices with frequencies f_1', f_2',and f_3' (and the response *neither* occurred with frequencies n_1, n_2, and n_3). As with the consensus data discussed earlier, the ranking index set aside the rarely occurring and uninformative *neither* responses, and was defined as follows:

$$\text{Ranking index} = \frac{f_1 + f_2 + f_3}{[f_1 + f_1'] + [f_2 + f_2'] + [f_3 + f_3']}$$

Since N = 100 subjects participated in each language group in the experiment reported here (so that f + f' + n = 100, in this instance), a logically equivalent formulation of the ranking index would be as follows:

$$\text{Ranking index} = \frac{f_1 + f_2 + f_3}{[100 - n_1] + [100 - n_2] + [100 - n_3]}$$

In brief, the ranking index pools data from the three competitions into which a vowel enters on a given scale, and is calculated as the ratio of the sum of evaluating choices favoring that vowel to the sum of all evaluating choices. These data will be presented in the form of line graphs, in which the ordering and spacing of stimuli on a scale reflect their relative ranking index values.

The ranking data were surveyed to see if the difference in clarity of response between salient and nonsalient scales observed in the consensus data might re-emerge in a different form: i.e., whether index values on the salient scales might span a wider area (from greatest to least) and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

prove more widely separated from each other than the values on the nonsalient scales. Rankings were also surveyed for similarities and differences between responses to the dimensions, and findings in this area have dictated how the dimensions will be grouped in all the subsequent analyses.

An important question that arises with the ranking data concerns a possible accidental character of the orderings, especially those of rank neighbors: are these merely the product of chance, or do they have a statistical basis? This question could be answered by integrating the two types of data, so that for each pair of rank neighbors it could also be seen whether there was consensus in differentiating between them. Accordingly, ranking data will be presented in conjunction with the relevant consensus data.

Index ranges. Using the ranking index values, it was also possible to obtain a third type of data. Index ranges were calculated as the difference between ranking index endpoint values (highest versus lowest). These range data allowed for an impressionistic comparison between any two scales, either within or across dimensions, one important focus being the relation between the salient and nonsalient pair member scales of a dimension.

Data types: summary and conclusion. Consensus distribution, rankings (subsequently integrated with consensus data) and index ranges were the three types of data extracted from the original frequency data. Cases of non-consensus were made an especial focus of investigation, as were the outcomes with the salient (cf. the non-salient) scale-pair members of the dimensions. Ranking data were scanned for similarities and differences in responses to the dimensions; these

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

findings structured the way the dimensions were grouped in interpretations of the data analysis. Finally, index range data provided the basis for impressionistic comparisons between the semantic scales, with particular focus on saliency contrasts within dimensions.

6.3 Additional approaches to interpreting the data

Stimulus vowels and the phonetic/phonemic backgrounds of the subjects. Chapters 7 to10 will treat in turn the data for each of the AE, Korean, Japanese and Spanish subject groups. Each chapter begins with a section describing the phonetic/phonemic background of the language group treated. In the absence of recorded data, these sections supply descriptions of each group's most likely versions of the point vowels, and of the central vowel, <ʌ>. Where the AE speakers are concerned, their renditions of the vowels may well not have been entirely uniform, and the ESL subjects were expected to have produced vowels intermediate between the English targets and their own..

In the comparison of articulatory accounts of the four target vowels in English and descriptions of their equivalents in the native languages of the ESL subjects, several problematic areas emerge. The first concerns <i> and <u>, since certain features considered part of the identity of English high vowels (Mackay, 1987) are not necessarily present in the high vowels of other languages. One such feature is lip gesture; the lip spreading described as intrinsic to English [iy] is not always present in the high front unrounded vowels of every language. Similarly, the feature of rounding is not intrinsic to high back vowels in all languages, but is considered an important part of English [uw]. Where speakers have first languages with minimal or no lip gesture for the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

high vowels (as is the case in Japanese), this fact may have played a part in the way they evaluated the high vowel stimuli. Also, the English high vowels have a feature variously intepreted as tenseness or diphthongization: this is definitely not the case for Spanish [i] and [u] (Schnitzer 1997), and descriptions of high vowel equivalents in Korean (Martin, 1992) and Japanese (Akamatsu, 1997) emphasize their monophthongal character. If AE models of the high vowel stimuli were attempted, then the presence of [y] or [w] offglides, or of a tense/diphthongized quality perceived as above-average duration, may have influenced the evaluations of some ESL subjects. Alternatively, ESL subjects may have given the high vowel stimuli a more native-speaking pronunciation, and this too could have affected evaluations.

A second area concerns the variability of subjects' realizations of <a> and <ʌ>. Descriptions of the tongue position for AE [a] suggest horizontal instability: we cannot specify for sure whether this vowel will be truly back, backed central or central (see Kurath and McDavid, 1961; Vance, 1987; Schnitzer, 1997). Moreover, descriptions of AE [ʌ] indicate that it is unstable in both height and anteriority (see the contrasting accounts of Ladefoged, 1982, and Mackay, 1987). A crucial further issue concerns the phonemic status, as well as the phonetic quality, of [ʌ]. While point vowel phonemes (albeit not realized in exactly the same ways from language to language) are common to all the languages of the participants, Japanese and Spanish have five-vowel systems with no phonemic equivalents to [ʌ]. Korean has a more elaborate vowel system, offering two possible candidates for an [ʌ] equivalent, but there is no inevitable choice, since authorities disagree about which is more likely (see Chapter 8, Section 8.1.1). Thus, each of the phonetic/phonemic background sections in the Korean, Japanese and Spanish chapters will present arguments about how the group in question is most likely to have 'solved the problem of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

<ʌ>'; i.e., how this vowel was most probably pronounced, and with which native phoneme it is most likely to have been associated.

The feature-based approach. The interpretation of the data was based on an examination of the features of the stimulus vowels, taking "features" in its widest sense to cover any characteristic which may serve to distinguish one vowel from another. As well as English-specific characteristics, these include those features of the ESL subjects' first-language vowels which may have been present in their pronunciations of the stimuli and which would have differentiated them from their AE counterparts, for example, the absence of rounding in the Japanese <u>-counterpart, [ɯ].

The context in which a stimulus was being judged was made up of two aspects: the semantic scale it was being evaluated on, and the other stimulus it was paired with. The combination of these two aspects was hypothesized to have a foregrounding effect on some feature or features of the stimulus vowel, while other features not felt to be phonosymbolically relevant would remain de-emphasized, in the background. On a scale X, in cases of non-consensus, the stimulus pairs were scanned for a shared X-connoting feature or features, i.e., ones which might plausibly be foregrounded by the semantic scale on which they were being judged. In cases of consensus, the stimulus vowel obtaining the majority vote was scanned for distinguishing X-connoting features, while its pair member was scanned for contrasting features.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 7

ANALYSIS AND INTERPRETATION OF THE AMERICAN ENGLISH DATA

7.1 The phonetic/phonological background of AE subjects

7.1.1 Articulatory details

In his account of the AE vowels, Mackay (1987: 72) describes [iy] as being "high front tense spread," [uw] as "high back tense and rounded," [a] as "low back unrounded" and [ʌ] as "lower-mid, back (or central) unrounded." He also notes that the high front and high back vowels "tend to be somewhat longer... and share... diphthongization" (Mackay 1987: 73). In his scheme, [ʌ] is located in the same area as the mid low back vowel [ɔ], distinguished from it only by lack of rounding. Ladefoged (1982: 198) places [ʌ] considerably lower (halfway between [ɔ] and [a], which he depicts as the lowest vowel) and centrally, almost directly below schwa: "Note that the symbol [ʌ]... is often used for a lowered mid-central vowel" (Ladefoged 1982: 203).

Another way of classifying the vowels refers to the position of the lower jaw rather than the tongue, and therefore contrasts their *closeness* or *openness* (see Brosnahan and Malmberg, 1970; Gimson, 1989; Jones, 1978; Laver, 1994). The vowels [iy] and [uw] with high tongue position are articulated with a close lower jaw, while [a], the vowel with low tongue position, is articulated with an open lower jaw. Jones (1978: 45) classifies [ʌ]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

as "half-open," but, depending on its variation in height, it might also be described as "nearly open." While "close" and "open" may appear to be redundant terms for "high" and "low," their usefulness is twofold. Firstly, they focus on the visual/proprioceptive aspect of the vowels in question (i.e., those differences which are perceptible to the interlocutor in the way that tongue positions are not, and which also recall associated proprioceptive sensations). And secondly, they suggest a direct link to differences in amplitude between vowels, given an implied presence or lack of labial/lingual obstruction, respectively, in the close and open jaw positions.

A detailed phonetic (rather than phonemic) description of the vowels likely to have been produced by the AE-speaking subjects in this experiment can be found in Kurath and McDavid (1961). This 40 year old study remains to my knowledge the authoritative source, not superseded by any more recent investigation; in general, Kurath and McDavid's descriptions should capture the pronunciations of the older participants and indicate variations that may have been present in the pronunciations of the younger subjects.[1] In the case of younger NYC English speakers, a general tendency towards fronting the vowels has also been noted, and if this were present in the speech of the younger AE-speaking participants, it may have changed their renditions of the non-front stimulus vowels.[2]

---

[1] Some two-thirds of the participants in this study were undergraduates under 30 years of age, but the remainder were graduate students and/or teachers in their mid-thirties, or older.

[2] This fronting tendency has been observed by Dr. Edward Bendix (personal communication). It could have rendered some younger subjects' pronunciation of [ʌ] more schwalike, and made [a] a central (rather than backed-central) vowel. In addition, it may have altered the 'long-tube' quality of [uw], depending on

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Kurath and McDavid's three charts of Metropolitan New York City speech show that these authors considered two words sufficient as examples of the high vowels and three for [ʌ]; however, they offered eight examples for [a].[3] This suggests a relative instability in [a], i.e., that it is more likely than other vowels to vary in different phonetic contexts. Kurath and McDavid's transcriptions and descriptions do indeed indicate the stability of the high vowels (i.e., that variation is minor only), and some tendency towards lengthening and monophthongization.[4] Surveying their transcriptions of NYC [a], we can see definite horizontal variability; there is variation between a low central unrounded and a low back unrounded vowel, depending on the word used. Overall, the impression given is that [a] in this dialect is a backed-central vowel. Their transcriptions for NYC [ʌ] also give an impression of its stability; its placement of on the vowel quadrilateral is different from both Ladefoged's and Mackay's, since these authors locate it as higher and more central than [ɔ], i.e., as a nonback mid vowel. In fact, Kurath and McDavid's [ʌ] seems very close to what was aimed at originally when the questionnaire was designed—the closest possible stressed vowel to schwa.

---

whether lip-protrusion/rounding were exaggerated in compensation for a more forward tongue articulation.

[3] Each chart gives a minimum of two words exemplifying any vowel phoneme of American English, together with informants' renditions (in narrow transcription) of the vowels as spoken in those words.

[4] All their high vowel transcriptions show length, but no offglides. My own observation is that in NYC speech, postvocalic glides are generally present on the high vowels, but that the *degree* of gliding is a feature of social class—the more working-class the speaker, the more fortis the glide. But in the case of this experiment, the presence or absence of the glide is less important than the most likely overall perceived length of the vowel; therefore, since either a glide or lengthening is likely to have been present on the these subjects' high vowels, an impression of their being comparatively longer than [ʌ] (and possibly than [a]) should have arisen in either case.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

7.1.2 Acoustic details

We now turn to the acoustic picture. Parameters were sought which would produce rankings for the four target vowels, so that these these could be subsequently compared with stimulus rankings derived from the behavioral data obtained in the experiment.[5] Four acoustic parameters were considered: formant locations and consequent main regions of energy concentration; overall amplitudes, derived from formant intensities; duration; and fundamental frequency values.

Formant locations and intensities. Table 7.1 below cites values for F_0, F_1, F_2 and F_3 drawn from the classic large-scale study of Peterson and Barney (1951).[6] As an aid to understanding each vowel's formant pattern, regions of acoustic energy concentration are indicated by links between formants which are narrowly separated.

---

[5] This approach follows the suggestion of Dr. Lawrence Raphael (personal communication).

[6] Male-voice values alone are given, partly in the interests of uniformity of acoustic comparison (since large-scale vowel formant data for female speakers could not be located for two of the language groups tested), but will suffice for the present purposes. There are some relational differences between AE male and female formant distributions; where two point vowel formants are relatively close together (e.g., [iy]'s F_2 and F_3), the relevant formants are closer for females than for males in every case. It is also true that [ʌ]'s very even male lower formant spacing is less even for females. However, to consider female formant values in addition to the male ones is to imply that there are grounds for thinking that not just acoustic quality but the actual *perception* of the vowels differs crucially between the sexes, and in turn this difference will be reflected in phonosymbolic evaluations.

Instead, it can be argued that production differences in vowels between the sexes are perceptually normalized by the listener through being translated into his or her vowel system (as are dialectal differences, or those between child and adult; see Joos, 1946: 61-62). Moreover, the main acoustic cue for distinguishing among vowels is the F_1- F_2 relation (Peterson and Barney, 1951), and whatever the age or gender of the speaker, these are apparently similar enough to be translatable. In addition, as will be seen in Chapter 8, data from Korean speakers showed a very high correlation between the sexes, in phonosymbolic responses.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 7.1: Formant frequencies (Hz) and intensities (dB, relative to F_0, in parentheses) for male speakers of American English

	F_0	F_1	F_2	F_3
[iy]	136	270 (-4)	2290 (-24)	3010 (-28)
[a]	124	730 (-1)	1090 (-5)	2440 (-28)
[uw]	141	300 (-3)	870 (-19)	2240 (-43)
[ʌ]	130	640 (-1)	1190 (-10)	2390 (-27)

Adapted from Peterson and Barney (1951: 183)

From Table 7.1, we see that [iy] has a distribution of acoustic energy with two widely separated peaks, the second of these yielding a notable feature of this vowel. None of its formants is especially intense (as compared to those of [a] or [ʌ]), but in the high frequencies it contrasts maximally with the less-intense [uw]. Despite the general lack of intensity of [iy]'s formants, it has been observed that its high-frequency concentration of acoustic energy, which could be described impressionistically as its "sharpness," has a carrying-power over long distances. Thus [iy] may be perceived as having a type of loudness.[7]

In contrast, the main concentration of acoustic energy for [a] is found in the middle of the spectrum, where its formants have relatively high intensity resulting from its jaw-

[7] This observation was made by Dr. John Roy (personal communication).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

openness and lack of oral obstruction, and one would therefore expect it to be perceived as the loudest of the four stimulus vowels. Here, we may contrast two possible types of perceived-intensity (loudness) features: the high-frequency carrying-power of [iy] and the overall amplitude of [a]. The acoustic energy for [uw] is concentrated at the lower end of the spectrum, and overall, its formants have the least acoustic intensity. Thus, we would expect [uw] to be perceived as the *softest* vowel.[8] In contrast with the point vowels, [ʌ] is best described as having an overall diffused distribution of acoustic energy; from its formant intensities, it is expected to be perceived as loud, rather than soft.

Intrinsic duration. This third acoustic parameter is included not only for its potential to rank the target vowels but because strong hints of its phonosymbolic importance are given by anomalies in the findings of Sapir (1929) and Newman (1933) in their investigations into phonosymbolic size.[9] Data were surveyed from the classic study of Peterson and Lehiste (1960), according to whom, [ʌ] can be distinguished as one of the "short nuclei," cf. [iy], [uw] and [a], among the "long nuclei" (Peterson and Lehiste, 1960: 701-703). These authors also measured acoustic duration in systematically sampled phonetic contexts, to arrive at intrinsic duration values. For the four stimulus vowels, these yield a ranking (least to greatest) with the following order: [ʌ], [iy], [uw],

---

[8] The impressionistic term "soft" will be used (as in normal speech) as the antonym for "loud." No reference to yieldingness or smoothness of texture (i.e.,"soft" as antonym for "hard") is implied.

[9] See Chapter 2 for a discussion of this point.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

[a].[10] There is a caveat here: in Peterson and Lehiste's study, although the number of tokens was large, the number of subjects was small.

Fundamental frequency. It can be argued that the features of intensity, of locus/absence of acoustic energy concentration and of intrinsic duration represent the three essential acoustic characteristics most likely to have been used in phonosymbolic attributions. However, a fourth characteristic has also been cited: in seeking to understand the symbolism of vowels, Tolman (1887, 1904) and Macdermott (1940) placed emphasis on *intrinsic pitch*, i.e., the perceptual correlate of inherent (non-intonational, non-expressive) differences in F_0. In Peterson and Barney's data (see Table 7.1), F_0 values for the four stimulus vowels support a ranking from lowest to highest intrinsic pitch as follows: [a], [ʌ], [iy], [uw]. Pitch differences here are not large, since the successive steps of this ranking differ by only 5 to 6 Hz. However, if Tolman and Macdermott are correct, we should find phonosymbolic contrasts in which at least the extremes of this order, differing by 17 Hz, and corresponding to stimuli <a> and <u>, are maximally opposed.[11]

Even so, intrinsic pitch seems likely to be one of the last acoustic features to be called into play to support semantic judgments because of the expressive fluctuation of F_0 in discourse. Any contextualization of a vowel that involves relativizing it (in an isolated

---

[10] These authors measured the effects of consonantal voicing, place and manner on vowels in *CVC* frames. In two out of their three tables of duration values, the ranking cited emerges. However, the vowels [uw] and [a], otherwise well distinguished, were reported to have the same length where where they were averaged over the utterances of 5 speakers on 70 minimal pairs.

[11] F_0 data for Korean vowels display the same ordering, i.e., maximally contrasting [a] and [uw] (see Yang, 1996: 251).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

word of more than one syllable, or in a phrase) will add intonational stress and so cause its F_0 to diverge from its inherent value. Thus, we might say that inherent F_0 (much more than F_1, F_2 or F_3) is especially unstable as a feature, hence unsuitable for phonosymbolic interpretation. But though vowels in discourse are bound to vary their F_0's, the F_0 relation between vowels in the pairing-context of an experiment like this one may still hold constant and serve to differentiate them. Thus, F_0 is included as a vowel feature (in the sense that feature is used here), available for phonosymbolic purposes.[12]

7.1.3 The vowel scales of two important predecessors

In this and the following chapters, reference will be made to the vowel scales of Tolman (1887, 1904) and Newman (1933). Unifying judgments of brilliance, size, mood and speed, Tolman's "tone-color" scale was arrived at introspectively. Newman worked with Sapir's (1929) experimental data and his own to produce a size scale; using experimental data of his own, he also produced a brilliance scale (for detailed accounts of both authors' work, see Chapter 2).

On several grounds, it may not be reasonable to expect a perfect correspondance between the behavioral rankings emerging in this experiment (whether for AE or ESL subjects) and these earlier vowel scales. The latter are quite extensive and more graded, perhaps more subtle in their distinctions: in their ground-breaking work, the early authors

---

[12] Recent experimental research on AE vowels suggests that F_0 does have its part to play in vowel identification: Halberstam (1998) found that where F_3 is not available, listeners will resort to F_0 to identify a vowel.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

investigated considerably more vowels than the 4 used in the present investigation, Tolman including 17 and Newman 9. At the same time, Tolman, but not Newman, evaluated [ʌ], an important part of the present experiment because of its centrality. Tolman's use of a *unified* cluster of attributes calls into question the precision of his intuitions. Certainly, the implication that the relation between brilliance and size is one which associates brightness with smallness (or darkness with largeness) may not be warranted. Indeed, the character of that relation is a focus of this experiment, and of Newman's (1933) investigations. It is not surprising Tolman and Newman conflict with each other in their vowel rankings.

Yet these earlier vowel rankings can serve as useful historical standards against which to compare those discovered in the present experiment. For this reason, illustrations of Tolman's tone-color and Newman's brilliance and size scales (taken from Chapter 2) are repeated below in Figure 7.1. In the figure, vowels occurring in the present experiment are emphasized.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

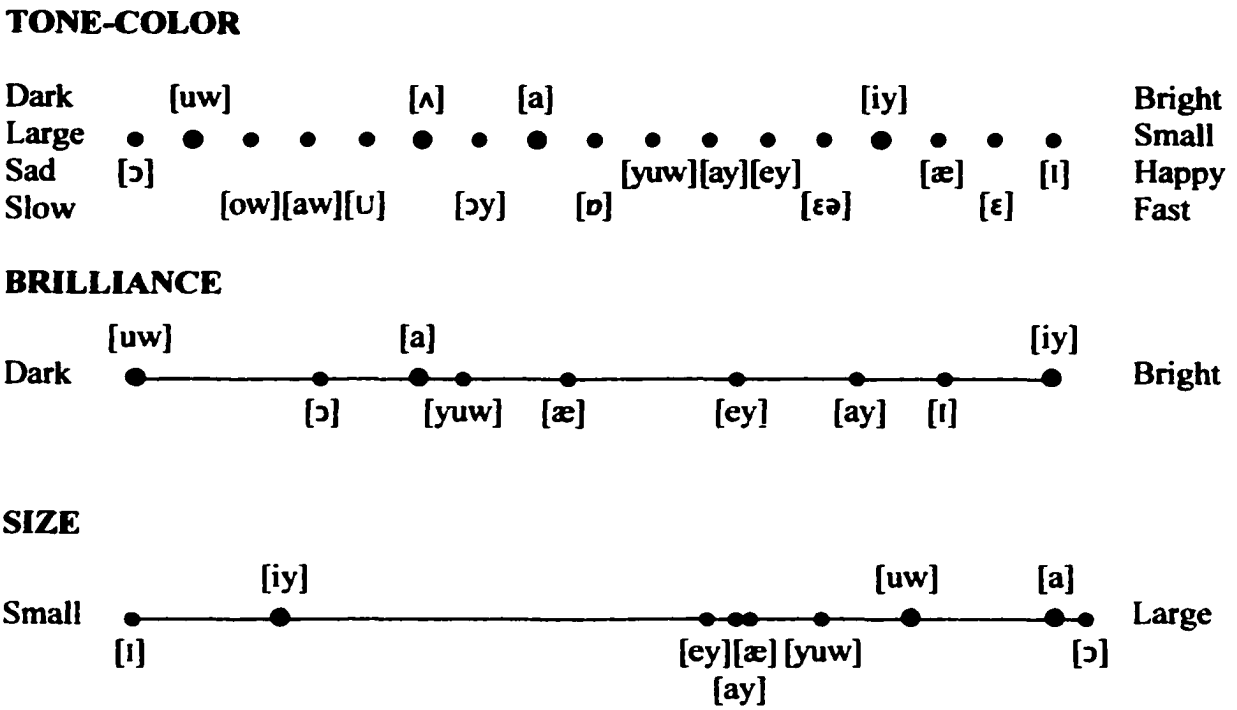


Figure 7.1, Illustration of early vowel scales. *Upper panel*: Tolman's (1887: 52) order for 17 American English vowels based on a composite of attributes; spacing is arbitrary. *Middle and lower panels:* Nine American English vowels scaled on brilliance and size dimensions; spacing represents Newman's (1933: 63) findings.

7. 2 Analysis of the AE data

Figure 7.2 below presents in summary form the stimulus rankings on six semantic scales, derived from the responses of the AE speakers. [13] The rankings shown in the figure are based on ranking indices (taking minimum and maximum values of 0 and 1 respectively), and are accurate in scale. These are integrated with consensus data, i.e., with the findings of statistical tests on all stimulus pairings; the more commonly-occurring case of

[13] See also Appendix 4, where Figure 7.2. is presented together with its equivalents for the Korean, Japanese and Spanish speakers.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

differentiation is assumed, and only non-differentiation is indicated. (See Chapter 6, Section 6.2 for details). In order to emphasize relations between antonymic scale pairs, rankings for the salient scales are ordered from least to greatest, reading left-to-right (e.g., least bright to brightest), while rankings for the non-salient scales are ordered greatest to least, reading right-to-left (e.g., darkest to least dark). Arrowheads on the scale lines are intended to remind the reader of this reversal of direction.

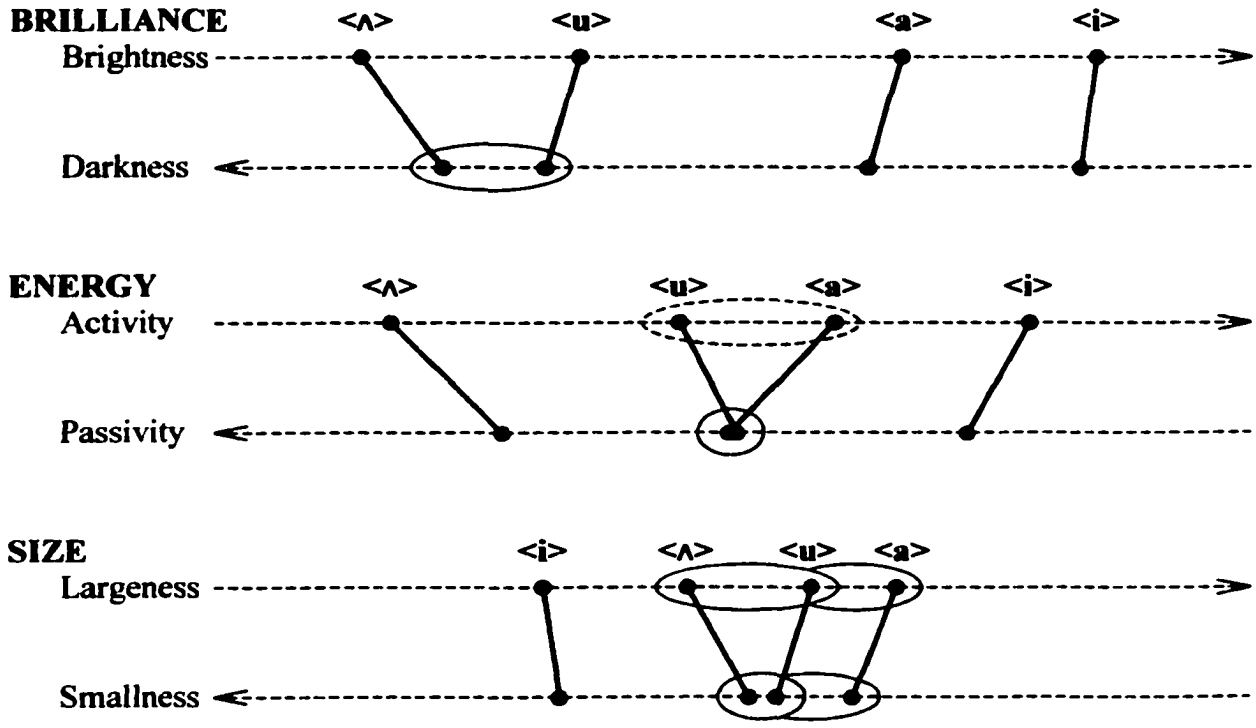


Figure 7.2, Data patterns (AE speakers). Ranking index values and consensus data for salient and non-salient sub-scales of three semantic dimensions. Ellipses surround stimulus pairs not differing by statistical test, $p>.05$; dashed outlines indicate instances of trend, $.10>p>.05$.

Inspection of Figure 7.2 reveals a broad similarity between the sets of stimulus rankings obtained on the brilliance and energy dimensions. All of the relevant scales (with one

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

minor exception)[14] pattern together. Thus, at one endpoint we see that the AE subjects judged <i> as brightest/least dark and most active/least passive, while at the other, they judged <ʌ> as the least bright/darkest and least active/most passive. The size dimension contrasts in displaying a different stimulus ranking. On the two size scales, although <i> is again ranked at an endpoint as least-large/smallest stimulus, the opposite endpoint is occupied by <a> (largest/least-small), and <ʌ> falls in a middle position, in contrast to its endpoint position on the brilliance and energy scales. Moreover, the stimulus pair <ʌ/u> was not differentiated on either size scale, whereas it was differentiated on three of the four brilliance/energy scales.

There are two ways of construing these similarities in ranking on the AE brilliance and energy scales, should these survive closer scrutiny. The first would be that some semantic archetype or *metadimension* (neither brilliance nor energy directly, but an attribute that would include both as subdimensions) was being accessed by subjects. The second would be that one of these attributes, either brilliance or energy, is semantically more basic and includes the other.[15] But any conclusions on what might be the true relation between the semantic dimensions (and thus the cause of ranking similarities) must wait until data from all four speaker groups have been analyzed. We note here only that the AE data provide a reasonable basis for categorization of the six scales into one grouping of four

---

[14] The exception arises with the stimuli <a> and <u> on the passivity scales.

[15] However, this second possibility is not supported by the metadimensional findings of the original authors of the semantic differential test, Osgood, Suci and Tannenbaum (1957: 63). These suggest that the relation between the attributes of brilliance and energy is not as straightforward as inclusivity of one by another; the authors observe that "in general scales representing this [activity] factor *are not as independent of other factors as would be desirable*" (emphasis added). Active/passive was found to be a "pivot" scale (i.e., the most highly loaded) on the metadimension of Activity. Dark/bright also had a positive loading for Activity, although more heavily loaded on the Evaluative dimension.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(the scales for the brilliance and energy dimensions) and another of two (the scales for the size dimension), and we will explore this possibility in what follows.

A second striking aspect of the AE data, implicit in the discussion above and explicit in the construction of Figure 7.2., is the symmetricality of the rankings for salient and nonsalient scales. However, it cannot be assumed that the basis for evaluations on any scale will necessarily be a simple reversal of those on its antonym. In general, the non-salient scales have narrower spacing than the salient scales; they may also have a greater extent of non-consensus.[16] Consequently, there are reasonable grounds for considering that antonymic questions were not simply experienced as positive/negative versions of each other, and this too will be explored in what follows.

A third and final issue is the importance of stimulus *contrasts* arising from the pairings of stimuli in the test items. Evaluation of a stimulus appears to have been affected not only by what scale it was judged on; which other stimulus it was being compared with seems also to have been crucial. Some stimulus pairings resulted in crisp, well-differentiated evaluations, with consensus across all six scales; in the AE data, Figure 7.2 reveals that this is the case for four pairs, <ʌ/a>, <ʌ/i>, <u/i> and <a/i>. On the other hand, the pairs <ʌ/u> and <u/a> were poorly differentiated, with non-consensus on three and four of the six scales, respectively.

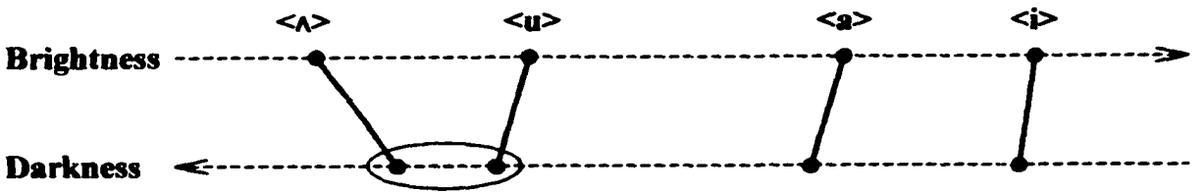
---

[16] In fact, the correlation between salience/non-salience and exent of consensus will be seen only as a mild tendency in the AE and Korean data, but is much more clearly exhibited in the Japanese and Spanish data.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

We now turn to a consideration of the data in detail, treating each semantic dimension in a separate subsection: 7.2.1 will discuss brilliance, 7.2.2, energy, and 7.2.3, size. A final subsection, 7.2.4, summarizes findings and presents a semantic profile of each of the stimulus vowels as evaluated by AE subjects. Those issues emerging in the discussion above, based on initial observations of the overall data pattern, will be explored as they arise in analyses and intepretations which scrutinize the findings more closely.

7.2.1 Brilliance dimension



Brightness

	<u>	<a>	<i>
<ʌ>	69 / 29	92 / 7	94 / 5
<u>		79 / 20	85 / 15
<a>			74 / 25

Darkness

	<u>	<a>	<i>
<ʌ>	43 / 55	12 / 87	10 / 89
<u>		27 / 72	12 / 86
<a>			27 / 71

Figure 7.3, Brilliance dimension data arrays (AE speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing significantly, $p>.05$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For the reader's convenience, Figure 7.3 above extracts the brilliance dimension data previously presented as part of Figure 7.2's overall summary of the data for AE speakers, and combines it with the frequency matrices[17] from which it derives. This permits a detailed consideration of the speakers' evaluations of the stimuli on brightness and darkness scales.

General remarks on the scales. To begin the examination of responses on the salient and non-salient scales in the lower panel, consider as an illustrative example the two left-most cells of the two frequency matrices (lower panel), those treating the stimulus pair <ʌ/u>. In each cell, the left-right alignment of the two frequency scores, together with their placement on lower and upper lines, indicates to which of the paired stimuli the cell entries correspond: here, <ʌ> is lower left and <u> is upper right. Thus we see that <u> achieved a majority on the salient brightness scale (gaining 69 of 29+69 choices), while <ʌ> achieved majority on the non-salient darkness scale (55 of 55+43 choices), suggesting symmetricality between the two sets of judgments. However, we also see that this symmetricality is only partial, since there is a substantial difference between the scales in *robustness* of responses evoked: the values given indicate that <u> is judged brighter than <ʌ> more robustly than <ʌ> is judged darker than <u>. Indeed, the salient/non-salient contrast in robustness very marked for the <ʌ/u>pair, since consensus in differentiating between the stimuli is reached only on the salient scale. (The shaded cell of the frequency matrix for the non-salient scale indicates non-consensus; this

[17] For the full AE frequency matrix figure (presenting stimulus-pair frequency totals for all three dimensions together), please see Appendix 5, which also presents corresponding data for the Korean, Japanese and Spanish speaker groups.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

corresponds to the shaded ellipse linking <ʌ> and <u> in the upper panel of the figure.) Although the <ʌ/u>pair is a particularly striking case, there is also a modest difference in robustness in the same direction (favoring the salient scale) in 4 of the remaining 5 brilliance comparisons.

The upper panel of Figure 7.3 shows that on both salient and nonsalient scales, *spacing* between the stimuli (i.e., separation among their ranking index values) is generally very good. The differentiation between paired stimuli is almost total, with only the above-mentioned case of non-consensus on <ʌ/u> in judgments of darkness. Moreover, there is similarity between the scales' *spacing configuration* (i.e., the way the four stimuli are positioned relative to each other), although there is a greater spread of index values for the brightness scale (from .139 to .844) than for darkness (from .166 to .780). To put this another way, there is a degree of "shrinkage" in the ranking index *range* for the non-salient scale relative to the salient (.710 points for brightness, versus .614 for darkness).

Interpretations of stimulus rankings. A striking aspect of these results for American English is how well the brightest/least-dark ranking for <i> agrees with the observations about this vowel in the literature.[18] For example, we see from Figure 7.1, on his integrated scale (whose attributes include brightness)Tolman (1887) ranks [iy] fourth out of his 17 vowels evaluated, while on his brightness scale, Newman (1933) ranks [iy] out of his 9 vowels at the brightest endpoint. The most plausible triggers of <i>'s brilliance evaluations are the tradionally-cited features of this vowel: its maximum anteriority and

[18] For additional remarks on the brilliance of [iy], see also Jespersen (1922), Chastaing (1962) and, more generally, Chapters 2 and 3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tongue height, and the acoustic correlate, a high-frequency acoustic energy concentration. In addition, <i>'s 'sharpness,' should be mentioned: this is a proposed loudness feature relating to the vowel's carrying-power. Furthermore, lip spreading on this vowel may be an important phonosymbolic factor; its non-linguistic analogue is the happiness gesture of smiling, and there is a strong association generally in this culture between brightness smiling, and being happy (brought out in common metaphors such as "a bright smile," "a bright face," "a bright glance"). We also see that in Figure 7.1, the attributes of happiness and brightness are clustered with each other on Tolman's (1887) integrated scale. Finally, we note that the smile gesture raises vowel frequencies in general (see Tartter 1980); thus, it may act as an enhancer of <i>'s high-frequency concentration.

On both scales, consensus is evoked for the rank-neighbor pair <i/a> and <a/u>. Given the good spacing of stimuli generally found on the brilliance scales, this suggests a perceived gradation in phonosymbolic brilliance between these pairs of stimulus vowels, rather than simple absence of brilliance.[19] As observed above, in Section 7.1.2, for [a], the main concentration of acoustic energy for is in the middle of the spectrum, and for [uw] in the lower part of the spectrum. Thus, the brilliance rankings for the point vowel stimuli appears to follow the descending order of locus of acoustic energy concentration. A second, proprioceptive, factor in the rankings of <a> and <u> can be found in the anteriority of the vowels relative to each other (i.e., in the articulatory correlate to their

---

[19] These findings agree with Tolman's (1904: 152) middle ranking of [a] on a seventeen-vowel scale better than Newman's (1933: 63) more extreme one of [a] as third-darkest on a nine-vowel scale. Both authors, however, agree on [uw] as definitely darker than [a]—second-darkest (Tolman); darkest (Newman)—which are in accord with the present findings.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

acoustic-concentration ordering). However far forward or backward <a> may have been articulated, it would still have been back of <i> and forward of <u>, that is, mid-ranked in anteriority. A third, linked factor, is <u>'s lip rounding. This has its bearing both on <u>'s tongue position (it enhances the backing tendency, i.e., lengthening this vowel's 'long-tube' quality) and its formant structure (it enhances the lowering of the formants: see Borden, Harris and Raphael 1994). Finally, a fourth factor polarizing <a> and <u> is <a>'s openness, which contrasts with <u>'s rounding: Whereas <a> has the least obstruction (labial or lingual) of the stimulus vowels and the highest amplitude, <u> has the greatest labial[20] obstruction of all the stimulus vowels and the lowest amplitude. We find here, in producing the brilliance-differentiation between these vowels, a visual, gestural component (labial rounding/constriction versus openness) linked to an acoustic-perceptual one (softness versus loudness).

Least bright and darkest on the AE brilliance scales was <ʌ>: What feature or features of this vowel could be responsible for its ranking? If we assume Kurath and McDavid's (1961) view of NYC <ʌ> hold for the vowel produced by the AE subjects in this experiment (though possibly colored by tendencies to velarization or backing),[21] <ʌ> is characterized by *centrality*. This could be described as a nexus of minus-features: lack of high or low, front or back tongue position, lack of jaw openness or closeness, of palatal or labial offglide, or of any noticeable lip gesture. The acoustic correlates of articulatory

---

[20] Lingually, <u> can be said to share the same degree of obstruction as the other high vowel stimulus, <i>.

[21] For evidence in support of the view that velarization and backing are features which correlate phonosymbolically with darkness, see the rankings of [uw] and [ɔ] on the vowel scales of Tolman (1904) and Newman (1933) in Section 7.1.3 above.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

centrality also demonstrate a minus quality: instead of a concentration of acoustic energy in one place in the spectrum, (as with [iy], [a] and [uw]) there is a diffusion, which could also be expressed as the absence of an acoustic-energy focal point. Masson (1952: 40, emphasis added) asserts that there is a phonosymbolic link between this diffusion and a perceptual muddiness: "...a natural unconscious relation in the human mind between the distance separating the formants 1 and 2 of a vowel and the luminescense of synesthetic colouring, and that *the even spacing over so wide an extent of the field of possible frequencies, of formants 1,2,3 in [central] vowels like [ə] and [ɜ] is directly connected with the muddiness of their synesthetic colors.*" In the range of possible associations, evidently, muddiness can be seen as close to darkness; it is easy to see how the perceptually muddy quality of <ʌ> could have been translated into nonbrightness/darkness in the context of judgments on the brilliance dimension. In sum, <ʌ> contrasts along various dimensions articulatorily and acoustically with the other three stimulus vowels; it is likely that its articulatory blandness and acoustic diffuseness/absence of concentration is responsible for its extreme ranking on the brilliance scales.

Another question raised by the brilliance dimension findings is about the collapse on the darkness measure of the pair<ʌ/u>: Which common feature(s) might have triggered their evaluation as dark stimuli? An initial survey seems to find more contrasting features than shared ones, since for AE speakers the most likely rendition of these vowels is likely to have differed in tongue height,[22] and perceived duration. Apart from the very general

[22] Evidence for their different degrees of jaw openness/tongue height can be found in their very different F_1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

characteristic of non-anteriority (shared also with <a>, which does *not* collapse with <u> or <ʌ>), the possibility of common articulatory features would arise only if these speakers had had tendencies towards backness and velarization of <ʌ>. On the other hand, there may be a common acoustic feature in a shared softness (perceived lack of intensity), compared to <a> and <i>; this assumes that the point made above about <i>'s 'sharpness' is correct (see Section 7.1.2).[23] It is also possible that subjects made an equation between <u>'s concentration of acoustic energy in the low frequencies and <ʌ>'s lack of any such concentration, i.e., that these two otherwise differentiating features were collapsed and interpreted as *equivalent* in the context of this scale. Finally, we should consider the logical possibility that the failure to differentiate between <u> and <ʌ> may have been an artefact of the experimental set-up: an orthographic association between the stimuli may have been provoked by the use of the letter *u* in the stimulus spelling *huh*.[24]

We must note, however, there *is* consensus in differentiating be tween <ʌ> and <u>on the brightness scale. In judgments on comparative brightness, we see the <ʌ/u> pair members pulled away from each other, with the higher, longer/tenser, more articulatorily extreme <u>, having some concentration of acoustic energy, being judged as brighter. But this

---

values (see Table 7.1).

[23] The assumption involves categorizing <i> as a loud rather than a soft stimulus. The hypothesis is that <i>'s lack of acoustic intensity due to the closeness of its articulation is outweighed by its sharpness. Among the stimulus vowels, formant data suggest that <ʌ> is middle-ranked in terms of intensity, but if it is seen as lacking both <i>'s sharpness and <a>'s unobstructed amplitude, <ʌ> should pattern with <u> rather than <a>, as would otherwise have been predicted.

[24] Orthographic *u* is ambiguous, in that it can represent both [uw] and [ʌ] in the English spelling system, as in *tub* and *tube*. See Chapter 6 for a full discussion of this point.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

discrepancy between the two sets of evaluations does not necessarily invalidate the phonosymbolic interpretations offered above. The semantic context afforded by the brightness question seems to highlight features which distinguish <ʌ> from <u>, while that afforded by the darkness question does not. The latter appears either to emphasize features that the two stimulus vowels have in common, or else fails to emphasize their distinguishing features, thus allowing shared features to dominate.[25]

7.2.2 Energy dimension

Figure 7.4 presents the data arising from judgments on the energy scales for AE speakers.

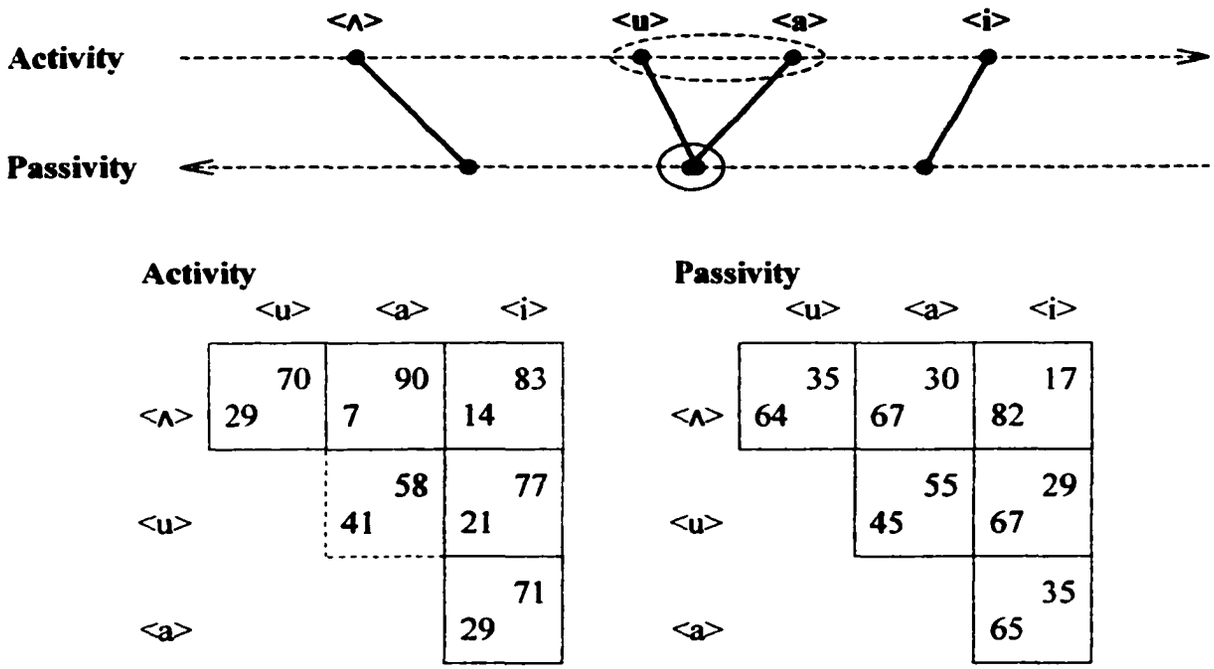


Figure 7.4, Energy dimension data arrays (AE speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates a stimulus pair not differing significantly, p>.05; dotted lines indicates trend, .10>p>.05.

[25] While it seems natural to think of different semantic contexts (brightness, darkness) emphasizing different aspects of a vowel's articulatory or acoustic realization, it is hard to imagine that this kind of influence extends to a consideration of a stimulus spelling. On these grounds, the orthographic account, included as a logical possibility, can be ruled out.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

General remarks on the scales. Both panels of the figure provide evidence that—just as for brilliance—the energy evaluations in general exhibit symmetricality between salient and non-salient scales, and at the same time, a difference in robustness of response. That is, with the notable exception of middle-ranked <u> and <a>, the stimuli are equivalently ordered from most active/least passive <i> through least active/most passive <ʌ> (see upper panel), and judgments of activity are more robust overall than those of passivity (see lower panel).The stimuli <u> and <a> are numerically distinct in their ranking index values for activity, but altogether collapsed for passivity and in either case produce the only occurrences of non-consensus when paired. Note also that cells for the <ʌ/a> pair in the frequency matrices exceptionally show no salient/non-salient reversal. These facts are reflected in the spread of ranking index values, which is markedly greater for the activity scale (.171 to .783, range .612) than for the passivity scale (.275 to .722, range .447). The size of this discrepancy suggests that subjects may have found it more difficult to evaluate stimuli in terms of passivity than activity; i.e., that their phonosymbolic intuitions were less accessible on the non-salient measure.

Energy compared with brilliance. The rankings of stimuli on the energy scales, apart from their weakness in the middle ranks, are quite similar to those on the brilliance scales, as noted earlier; this is particularly so for the salient scales. We have seen also that on both dimensions, the salient scales support more robust responses than the non-salient. This being so, it seems useful to ask whether the dimensions are also similar in their ranking index ranges. Table 7.2 gathers together these data.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 7.2: Ranking index ranges for salient versus non-salient scales of two semantic dimensions

	Salient	Non-Salient
Brilliance	.710	.614
Energy	.612	.447

Table 7. 2 suggests exactly the reduction in range evident within dimensions also holds between dimensions, brilliance affording a wider spread than energy. Indeed, the ranking index range for the *nonsalient* brilliance scale (darkness) is not greater that for the *salient* energy scale (activity), which suggests that dimension and saliency act jointly to determine the robustness of responses. To the extent that range indicates ease of phonosymbolic judgments,[26] then this suggests that brightness judgments may have been the easiest type for AE speakers to make. Darkness and activity were equivalently easy (but less than brightness), while passivity seems to have been the most difficult semantic scale to apply phonosymbolically.

Differences like those summarized above suggest that despite an overall similarity, brightness and activity (or darkness and passivity) should not be treated as synonymous for the purposes of phonosymbolic judgments. This caution isfurther strengthened by the observation that the energy scales show a different configuration from the brilliance scales, notably in the spacings of <a>, <u> and <ʌ>. On both brilliance scales, <u> is well separated from its rank neighbor, <a>; these exhibit non-consensus on both energy

[26] Since width of range correlates with good spacing, it can be claimed that the spacing displayed on a scale

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

scales. Conversely, on both the energy scales, <ʌ> is well separated from <u>; these exhibit non-consensus on the darkness scale (though they are differentiated on brightness). An almost complementary distribution of non-differentiation of this kind is important evidence that the grouping of brilliance and energy should not be construed as inclusivity of one dimension by another.

Interpretations of stimulus rankings. Since neither Tolman nor Newman presented judgments of activity/passivity on their scales, we cannot directly compare the rankings obtained here with the precedents set by earlier studies. However, Tolman's composite-attribute scale included the polar opposites bright/dark, happy/sad and fast/slow: the semantic field covered is thus broad enough that there are some grounds for associating "active" with the bright-happy-fast cluster and "passive" with the dark-sad-slow one, most especially with the speed antonyms. But comparison of Tolman's rankings with the present ones indicates considerable disagreement on all except for the most-active/least-passive ranking of <i>, Tolman's [iy] being comparably ranked, well towards the "fast" end of his scale, in fourth place. However, five rank places separate Tolman's [uw] from [a], while the stimulus pair <u/a> are collapsed on the present energy scale; in addition, Tolman ranks [ʌ] in the lower middle range, while here it is ranked at an endpoint, as least-active/most passive

Many of the interpretations of the stimulus rankings given in the treatment of brilliance (Section 7.2.1) are relevant here. The features of <i> that have already been proposed as likely triggers of brightness attributions are its high-frequency acoustic energy

---

is an indicator of how accessible a phonosymbolic measure subjects found it to be.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

concentration and its consequent 'sharpness' (a form of perceived loudness), and its high front tongue position. In addition, there are possible cultural factors: in New York City, the place where the AE subjects were recruited from we describe someone (approvingly) as being a "high energy" person, and it thus may be no coincidence that that <i> has the vowel which shows the greatest concentration of acoustic energy in the high frequencies and that it is judged the most active stimulus. Lip spreading may again be an important feature: not only is it a high-frequency acoustic energy enhancer, (Tartter, 1980), but its non-linguistic counterpart is the smile (a happiness gesture); there is a strong association generally in this culture between being active and being happy.

In addition, when we consider the maximum contrast of <ʌ> and <i> on the energy scales, another possible trigger may be phonological-phonotactic, an equation of the free/checked vocalic distinction with active/passive. In English, there is an opposition between the tense, diphthongized character of [iy] in a system where these features mean syllabic independence—the ability to be both nucleus *and* coda—and the lax, nondiphthongized nature of [ʌ], where laxness means syllabic incompleteness, the need for a following consonant to close the syllable in which this vowel occurs. Further evidence that AE subjects may indeed have intuitively equated free vowels with "active" and checked vowels with "passive" is found in the consensus in differentiating <u> (another tense, diphthongized free vowel) from its rank neighbor <ʌ> on both energy scales. It can be argued that the free/checked distinction is activated by energy judgments (which thus phonosymbolically separate <u> from <ʌ>) yet is irrelevant to those of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

brilliance.[27] Here, as elsewhere, the phonosymbolic applicability of a feature must depend on the context (i.e., the semantic scale).

As with <i>, many of the observations made earlier about <ʌ> are also relevant here: it can be claimed that <ʌ>'s articulatory centrality and acoustic diffusedness are as applicable phonosymbolically to (lack of) activity/passivity as they are to (lack of) brightness/darkness. If <i> could be described as a high-energy vowel, having a decided 'personality,' then <ʌ> could be characterized as having a *nebbishy* quality, a lack of specificity or anonymity. Two further connotations of anonymity are provided by a discourse-function of <ʌ> and by a phonotactic function: in American English, <ʌ> is the normal hesitation vowel, and it is also the normal epenthetic vowel (grading into schwa, depending on its degree of stress).[28] The first function associates <ʌ> with undecidedness, tentativeness, and the second with insertion of vocalic spacing (as between consonants in a cluster) without the distraction of place features; both seem to depend on <ʌ>'s lack of vocalic specificity. I would suggest that it is this core characteristic of <ʌ>'s anonymousness that lends itself to being interpreted as passive.

On both energy scales, <a> and <u> are separated from the extreme-ranked <i> and <ʌ>, but subjects did not differentiate between these mid-ranked stimuli on either scale. The crucial question is thus the cause of the collapse. One possibility is that different

---

[27] The collapse of <ʌ> and <u> on darkness gives evidence that the free-checked distinction is not relevant in brilliance judgments.

[28] Examples of this type of epenthesis would be emphatic [pʌl-liyz] (*please*) and [kʌr-rayst] (*Christ*), with secondary stress on the first syllable.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

features of each stimulus could have been judged vowel-internally as conflicting with each other i.e., each stimulus was felt to have both 'active' and 'passive' features whose effects neutralized each other. A similar possibility would be if features of both vowels (independently of each other) were felt to be ambiguous in their energy-connotations. This would have neutralized the phonosymbolic effect of each of the two stimuli, resulting in a shared middle ranking for both, though for two different sets of reasons. Another possibility is that rankings were the result of a shared, foregrounded feature or features (as has been argued for the collapse of <ʌ> and <u> on the darkness scale). We consider these possibilities in turn.

Vowel-internal feature-conflicts or feature-ambiguities of <u> and <a>. Could two independent, entirely separate vowel-internal conflicts, or ambiguities particular to each vowel, have resulted in similar energy-evaluations for <a> and <u>? Features arguably connoting <a>'s activeness include the following: non-low concentration of acoustic energy, non-backness, and acoustic intensity. Features potentially symbolizing <a>'s passiveness (or neutralizing its activeness) are non-high concentration of acoustic energy, non-frontness, and low intrinsic F_0. Both lists clearly suggest that since locus of acoustic energy and anteriority are not extreme for <a>, (for example, in the way these are extreme in the case of <i>), they may be capable of interpretation in either direction. Furthermore, <a>'s 'active' acoustic intensity and 'passive' low F_0 are both reflexes of the same articulatory feature: jaw openness/low tongue position. Thus it can be argued that <a>'s energy-features are ambiguous.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

On the other hand, it is hard to find features of <u> which could be reasonably claimed to connote activeness; impressionistic support for this view can be found on Tolman's integrated vowel scale where [uw] is ranked as second-darkest/saddest/slowest out of the 17 vowels considered. A sketchy argument could be made for <u>'s tenseness/diphthongization as associating it with the most-active <i>, but <u>'s only really plausible claimant as an activeness-feature is its high intrinsic F_0. On the other hand, a strong argument can be made that <u>'s passiveness might have been connoted by four of its features. These are backness and lip rounding,[29] which involves a pouting lip gesture (i.e., a sad one) in opposition to the (smiling, happy) lip spreading of the active <i>. In addition, <u>'s perceived softness and its locus of concentration of acoustic energy in the low frequencies can be argued as likely to connote passivity. In sum, whereas a reasonable argument can be made for certain of <a>'s features as ambiguous with regards to energy, <u> appears to have many more unambiguously passiveness-connoting features. Thus the <u/a> collapse cannot be explained as the product of two independent cases of feature-conflict or feature-ambiguity.

Shared energy-connoting features of <a> and <u>. The search here can be narrowed down to a feature or features likely to have foregrounded by the energy measures, but not by the brilliance measures (on which <a> and <u> were well-differentiated). Moreover, any claimant must plausibly rank neither <a> nor <u> at either extreme of the energy scales.[30] The most plausible candidate as a collapse-promoting feature is the same one

[29] Eight out of the 9 vowels ranked at the darkest/saddest/slowest extreme of Tolman's scale are back vowels, the exception being [ʌ]. Moreover, the lowest-ranked 5 all have lip rounding.

[30] If long intrinsic duration could connote passivity (which is not implausible as a kind of secondary

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

proposed in previous discussion of the brilliance scales: gradation/absence of locus of acoustic energy concentration. To recapitulate, it was suggested that on the brightness scale, this feature accounted for phonosymbolic differentiation between all four stimuli (in the ordering least-bright to brightest) as follows:

4) *absence* <ʌ>; 3) *low* <u>; 2) *medium* <a>; 1) *high* <i>.

In order also to account for the <ʌ/u> collapse on the darkness scale, it was proposed that the two lowest acoustic gradations had been blurred and reclassified as one (a *feature equivalence*), producing a three-step ranking (in the ordering darkest to least-dark) as follows: 3) *low/absence* <ʌ/u>; 2) *medium* <a>; 1) *high* <i>.

Similarly, another reclassification of acoustic gradations could account for the <u/a> collapse on the energy scales. While remaining sensitive to the same acoustic criterion, the focus changes somewhat: here, the greatest polarity is between the acoustically-diffused muddiness of <ʌ> and the sharpness of high-frequency-concentrated <i>, with the middle-ranked stimuli differentiated from both, but not from each other in the new semantic context. Thus the following ranking is produced (in the ordering least-active to most-active and most-passive to least-passive):

3) *absence* <ʌ>; 2) *medium/low* <u/a>; 1) *high* <i>.

If this interpretation is correct, then the brightness scale was thus sensitive to all four acoustic gradations, while the darkness and energy scales produced different feature-equivalences. A feature-equivalence of middle and low concentrations of acoustic energy would therefore be responsible for the middle rankings and collapse of <a> and <u> on

---

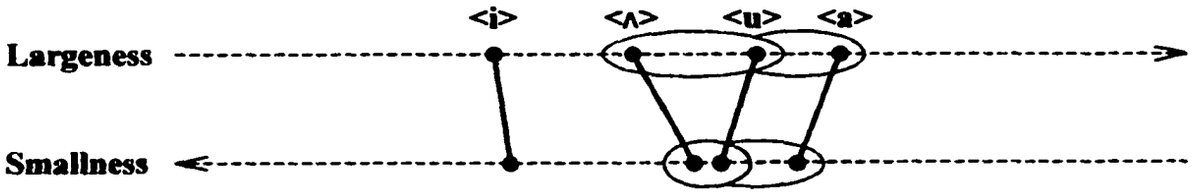
quality) then this might have been a passivity feature of <a>'s which was shared with <u>. But if duration had been a relevant trigger, it would have ranked both vowels at one extreme of the scales, and <ʌ>, the only lax stimulus vowel and the one with the shortest intrinsic duration, at the other. The facts therefore rule out this possibility: <ʌ>, and not <u> or <a>, was judged least active/most passive.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the energy scales.[31] To conclude, this suggests that locus of acoustic energy concentration could have been the basic triggering feature; thus, with certain context-sensitive modifications, it could be primarily responsible for the ranking agreements on both energy and brilliance scales.

7.2.3 Size dimension

Figure 7.5 presents the data arising from judgments on the size scales for AE speakers.



Largeness

	<u>	<a>	<i>
<ʌ>	54 / 46	76 / 22	32 / 67
<u>		54 / 43	26 / 72
<a>			36 / 64

Smallness

	<u>	<a>	<i>
<ʌ>	47 / 52	38 / 59	67 / 32
<u>		44 / 55	64 / 34
<a>			67 / 32

Figure 7.5, Size dimension data arrays (AE speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing significantly, p>.05.

---

[31] A similar kind of reclassification with a collapse could also have operated for the articulatory feature of anteriority: while <a> and <u> can be described respectively as middle and back vowels, both could also have been classified more broadly as nonanterior (i.e., a feature equivalence of non-front and back). However, <ʌ> is also non-anterior, and if a collapse had occurred due to this kind of broadening of classificatory boundaries, it should have been three-way and included <ʌ>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

General remarks on the scales. Figure 7.5 indicates a restricted spread of values on both scales, with somewhat greater compression on the nonsalient scale. However, both scales' index ranges are narrow, and there is only a small difference between them (a discrepancy of .059 index points). As observed with the other dimensions, the non-salient scale shows a diminution in robustness of response, but in the case of size, this is minor: the lower panel of Figure 7.5 indicates that three cells modestly favor the salient scale, one the non-salient scale and two are the same.

Both scales show the same spacing configuration: one stimulus, <i>, is evaluated as least-large/small, and there is what could be described as a continuum of the other three 'larger' stimuli, with non-consensus occurring between the two rank neighbor pairs on both scales. (In this continuum, however, the pattern is of two collapsed pairs, <u/a> and <ʌ/u>, and not of a three-way collapse; i.e., there is still consensus in differentiating between <ʌ> and <a>).

Size compared with brilliance and energy. One explanation for the lack of difference in robustness between salient and non-salient size scales may be found in comparing the stimulus configurations shown for the three dimensions in Figure 7.2 . We see that size judgments have moved towards the middle of the range, and that they show substantially more non-consensus (for both scales, 4 out of the 12 stimulus pairs). In effect, what we see is a three-step ranking, as follows: <i>, <ʌ>, <a>, with <u> *indeterminate* between <ʌ> and <a>, i.e., participapting in collapses with both of its rank neighbors.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In addition, of the three dimensions, both sub-scales of size show the narrowest spread of index values. As an aid in comparison, Table 7.4 below gives the index value ranges for the six AE scales in order from widest to narrowest.

Table 7.3: Ranking index ranges and range discrepancies, for salient versus non-salient sub-scales of three semantic dimensions

	Salient	Non-Salient	Discrepancy
Brilliance	.710	.614	.096
Energy	.612	.447	.165
Size	.342	.283	.059

We also see from Table 7.4 that size has the smallest index range discrepancy of all three dimensions. And lastly, as observed in the opening remarks of this section, the size rankings are notably different from those for brilliance and energy.

Thus, by index range, robustness and stimulus spacing criteria, both size scales appear to have been the most difficult of all six on which the AE speakers were asked to judge stimuli. This is a very surprising result. After all, the study of phonetic symbolism in this country began with Sapir's (1927) observation that it was easy for speakers to make comparative phonosymbolic size-judgments of a minimal pair of nonsense words. The implicit assumption was that for the purposes of investigating phonetic symbolism, size was a highly accessible semantic dimension, Sapir's later phonosymbolic size-study (1929) being based on that assumption. Yet here we have evidence that, of the three dimensions tested, the AE subjects found size the least accessible for judgments.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

What could have caused these surprising differences between the size results and those of brilliance and energy? We have noted that on the latter two dimensions, the ranking of all the stimuli could be correlated with gradations and reclassifications of a single triggering feature: location of concentration of acoustic energy. Other features, (such as anteriority or lip gestures) were proposed as providing additional influences, but these were not seen to conflict with the effects of this dominant trigger. If, in contrast, the size scale results are less clear, then one possibility is that there may not have been a dominant size-trigger to unify judgments in the way that was suggested for the brilliance and energy scales. An alternative possibility is that even if a dominant size-connoting feature were operating across all four stimuli, additional triggering features[32] may have been foregrounded by the pairing of different stimulus vowels. In either case, subjects would have been more confused in making their judgments.

The reader is reminded that although the rankings present a picture of all four stimuli relative to each other on a single scale, implying a unification of evaluating criteria, what the subjects experienced in each item of the questionnaire was an isolated comparison between two stimuli. On the same scale, then, there was nothing to prevent the subject from, e.g., comparing <i> with <u> phonosymbolically on the basis of one set of shared or contrasting features, <i> with <a> on the basis of another, <a> with <ʌ> on a still different basis, and so on. Thus, conflicts between two competing triggers, or between a dominant and additional triggers, might have been responsible for the lesser definition of

[32] The possibility of conflicting size triggers was originally suggested by a crucial observation of Sapir's (1929) on an anomaly in his size rankings, and by a similar observation of Newman's (1933), in his continuation of Sapir's study. For discussion of this point, see Chapter 2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the outcome for size (more restricted ranking index ranges, more collapses of stimuli) compared to those for the other two dimensions.

Potential size-triggering features. At this point, it is therefore worth reviewing those features with potential to act as size triggers and comparing the different rankings which would be generated by each with the actual rankings obtained in the data. Table 7.4 below gives five candidates.

Table 7.4: Observed stimulus ranking for size dimension (order being specified as least size = 1 through most size = 4), compared with rankings predicted by potential size-triggering features

Observed Ranking	**1** <i>	**2** <ʌ>	**3** <u>	**4** <a>
Articulatory spaciousness	1	2	3 *(or 4)*	4 *(or 3)*
Perceived duration	2	1	3	4
Perceived intensity	2	3	1	4
Intrinsic pitch	2	3	1	4
Energy concentration locus	1	4	3	2

First, the obvious (and traditional) articulatory candidate is the feature of oral spaciousness,[33] which has a clearly iconic relation to the size dimension. This can be subdivided into two dimensions: the vertical (as in <a>'s low tongue position/openness)

[33] For corroboration of this view (and its opposite, that the non-spacious [i] is small) see Paget (1927, 1930), who elaborates on the gestural aspect of phonetic symbolism, Jespersen (1933), and of course, Sapir (1927, 1929) and Newman (1933). See also, more generally, Chapters 2 and 3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and the horizontal (as in <u>'s 'long tube' quality, a combination of lip rounding with high back tongue position).[34] Second and third are the most likely acoustic triggers, perceived duration (length) and perceived intensity (loudness). Like articulatory spaciousness, both have an iconic relation to size, but they are of different linguistic types; they may also appeal to different levels of the phonosymbolic response. [35] Fourth and fifth are less likely acoustic size triggers, intrinsic pitch (the perceptual correlate of F_0), and locus of main acoustic energy concentration. In these cases, the iconic relation between size and the feature is less clear, though an argument can be made for high intrinsic pitch connoting smallness and low intrinsic pitch the reverse, and similarly for high frequency locus of concentration connoting smallness and low frequency locus the reverse.

Interpretation of stimulus rankings. In examining Table 7.5, it is clear that the size rankings of all four stimuli can be explained in terms of articulatory spaciousness: the non-spacious <i> is least-large/smallest, the neutrally spacious <ʌ> is ranked next to it; towards the large/non-small ends of the scales, we find the horizontally spacious <u>, while largest/least-small is the vertically spacious <a>. However, the spaciousness feature alone will not explain the shrinkage in index value range for size (compared to the

---

[34] Interestingly, the vowel which combines both types of articulatory spaciousness, [ɔ], is ranked largest on Tolman's (1887) integrated scale and Newman's (1933) size scale (see Figure 7.1).

[35] Vowel duration in English has a linguistic, phonemic significance (as a reflex of the tense/diphthongized feature), though it can also have a paralinguistic, expressive meaning; in contrast, vowel amplitude is a purely paralinguistic, expressive feature. Length, like articulatory spaciousness, implies a one-for-one relation between a linguistic quality and a real-world attribute; that is, both features make connections in the domain of phonetic symbolism. On the other hand, loudness suggets a non- (or pre-) linguistic, sound-symbolic connection (large beings and objects tend to make louder sounds than do small ones).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ranges of the other dimensions). Nor will it account for the notably poorer stimulus spacing on this dimension, which includes the collapses of two pairs of rank neighbors and the indeterminacy of <u>. A number of questions therefore remain:

1) Why is <i> the only stimulus on either scale not to collapse with its rank neigbor, <ʌ>?

2) Which features were likely to have been responsible for the collapse of <ʌ> and <u> on both scales? Are these the same ones as suggested for the darkness scale?

3) Why is there consensus in differentiating between <ʌ> and <a>, the endpoints of the three-stimulus large/non-small continuum?

4) Which features were likely to have been responsible for the collapse of <a> and <u> on both scales? Are these the same ones as suggested for the energy scales?

5) Why is the ordering of <u> indeterminate with respect to <ʌ> and <a>?

The rankings of <i> and <ʌ>: Because <i> is judged as less-large/smaller than <ʌ>, this argues that duration-related features were not the determining triggers for size judments in the comparison of these two stimuli.[36] Instead, it appears that other features of <i>, such as nonspaciousness, dominated durational features[37] to produce impressions of smallness. There are two explanations for consensus in differentiating <ʌ> from its rank

[36] See Section 7.1 above: by phonemic (tense/lax, diphthongized/non-diphthonigized) and intrinsic durational criteria, <i> should have been perceived as longer than <ʌ>.

[37] Intuitively, it seems that perceived vowel duration (produced by diphthongizing, lengthening or some other articulatory manifestation of tenseness) should have as good an iconic claim to phonosymbolic largeness as articulatory spaciousness. But from Sapir's and Newman's observations, it appears that duration is likely to dominate placement (i.e., horizontal or vertical spaciousness) only where the placements of the competing vowels are similar (as in the Newman's noted case of [iy] versus [ɪ]; see Newman 1933: 68).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

neighbor, <i>; the simplest is that <i> is classified by subjects on the basis of spaciousness as the only anterior (i.e., non-spacious) stimulus vowel, and in this respect, was felt to be as well differentiated from the central <ʌ> as from <a> or <u>. A more complex explanation is that though<ʌ>'s second-smallest ranking is mainly due to its anteriority, other features contribute to its placement. Its shortness[38] (lesser perceived duration) ranks it as less large/smaller than either <u> or <a>, most notably differentiating it from <a>, the longest stimulus vowel; this pushes it in the direction of <i>. But in addition, as emphasized in the size and brilliance analyses, <ʌ> maximally contrasts with <i> in terms of locus of concentration of acoustic energy. When the two stimuli were compared on the size scale, it is possible that this feature continued to be foregrounded, ensuring a polarization which distanced <i> from <ʌ>.

The collapse of <ʌ> and <u>. The important thing here is not just the collapse but the middle ranking of these two stimuli: triggers have to be sought which will explain both factors. The proposed shared features of <ʌ> and <u> determining their collapse on the darkness scale (shared softness, and a feature equivalence of low frequency acoustic energy concentration with absence of concentration) cannot be seen as responsible in this case because they are not reconcilable with a middle ranking. If these features were operating iconically, then softness would have produced a non-large/small ranking, as would have the (collapsed) low/absence frequency feature. As Table 7.5 makes clear, ranking by perceived duration would not explain the lack of differentiation between these stimuli either, since we would have the second-longest vowel, <u>, collapsed with the

[38] See Peterson and Lehiste, (1960); [ʌ] has the shortest intrinsic duration of all the stimulus vowels.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

shortest one ,<ʌ>. Similarly, Table 7.5 indicates that ranking by intrinsic pitch criteria would not substantiate a collapse; we would have the vowel with the highest intrinsic pitch, <u>, not differentiated from <ʌ>, which has the second-lowest. In sum, it appears that whichever feature(s) might have determined this collapse it, or they, are not acoustic. Instead, I would argue that the collapse is due to ranking in terms of articulatory spaciousness: the stimuli are seen to share horizontal spaciousness (non-anteriority) together with lack of vertical spaciousness (non-openness). If in addition <ʌ> was velarized, this may emphasized the shared non-openness, i.e., both vowels shared a constricting of the oral cavity at the velum, in contrast to the unconstricted quality of <a>.

The differentiation of <ʌ> and <a>. If above interpretation is correct, then one feature differentiates <ʌ> from <a>: articulatory spaciousness, but specifically the vertical dimension, i.e., <ʌ>'s half-openness versus <a>'s openness. If anteriority (i.e., the horizontal dimension) were the only criterion of spaciousness, we would expect a collapse between <ʌ> and <a>, since both are in the backed-central area, instead of the separation which we find. Table 7.5 indicates that a second differentiating feature is perceived duration: the contrast of, respectively, the shortest versus the longest vowel.

The collapse of <u> and <a>. Judged respectively largest and second-largest stimuli (their rankings here agree with both Tolman's and Newman's), one explanation for the collapse of these stimuli is that a form of (sub)feature-equivalence was operating, in which <a>'s vertical spaciousness was equated with <u>'s horizontal spaciousness, and thus both were seen as of equivalent articulatory largeness. Examination of Table 7.5

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

suggests that another type of size criterion, perceived duration, may also have had its influence on the collapse of <u> and <a>. Another shared feature of these vowels is non-anteriority, but if this alone were triggering size evaluations, we would expect a three-way collapse with the non-anterior <ʌ>, which was not the case. An argument along the same lines would rule out <a>'s and <u>'s common non-high locus of acoustic energy as responsible for their non-differentiation, since this should also have led to a collapse with <ʌ>. Finally, perceived loudness can be ruled out as a collapsing influence, since by this criterion, <u> and <a> would be ranked at opposite end-points (see again Table 7.5).

The feature-harmony of <a>, feature-conflict of <u>, and <u>'s indeterminacy The final column of Table 7.5 indicates that <a> is ordered at one extreme for four out of five potential size triggers; it is ranked highest for articulatory (vertical) spaciousness, long perceived duration, perceived intensity and intrinsic pitch. Thus, <a> can be claimed to show harmony in its size-connoting features. In contrast, although <u> is ranked second-highest for articulatory (horizontal) spaciousness and perceived duration, it is ranked lowest for perceived intensity and intrinsic pitch. In terms of size, <u> can therefore be claimed to show conflict in its triggering features, and this may be the reason for its indeterminacy. Its perceived duration and spaciousness associate it phonosymbolically with the large/non-small <a>, while its lack of perceived intensity, and possibly also its non-openness, move it towards a middle rank and to <ʌ>. In addition, its high intrinsic pitch may have pulled it towards <i>, at the nonlarge/small ends of the scales.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

7.2.4 Concluding remarks on the AE data and a phonosymbolic profile of the stimulus vowels according to the AE subjects.

Salient versus nonsalient robustness of response. The first point to be made is that there is a general tendency, separately noted so far with each dimension, for the salient scales to yield greater robustness in responses than the non-salient scales. Figure 7.6, the scatterplot below, documents this general tendency.

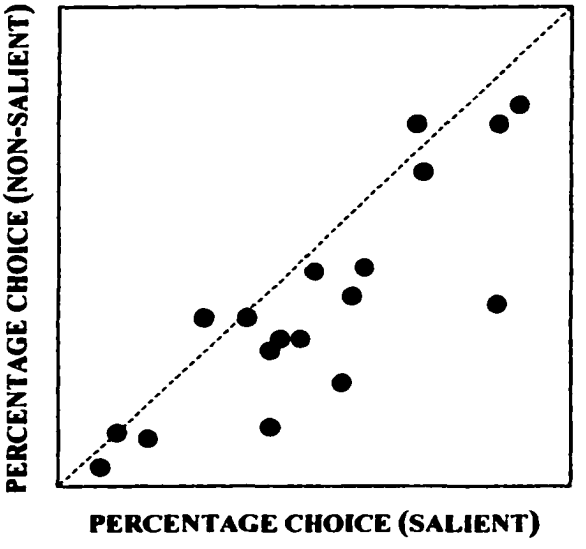


Figure 7.6, Preference strength in antonymic scales (AE speakers). Percentage of elective choices ('neither' responses omitted) favoring preferred stimuli for each of 6 stimulus pairs evaluated on salient vs. non-salient sub-scales of 3 dimensions. Axis values run 50 to 100%.

Here, dimension by dimension, corresponding cells of salient and non-salient scales in the frequency matrices were examined, and the preferred stimulus in each comparison, was plotted as a proportion of the elective choices which it received. These graphical representations of comparisons were made without regard to stimulus-pair or dimension. The horizontal axis represents the salient scale and the vertical the non-salient scale, while the diagonal indicates the expected average position if there were no salient-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

nonsalient distinction. However, the greater tendency for robustness in the salient scales is reflected in the fact that values tend to fall below the diagonal (i.e., in the salient half of the of the field); we see only two instances of values appearing above the diagonal. The distribution here clearly represents the tendency, reported from surveying the frequency matrices, that overall salient scale responses are sharper in distinctions than non-salient.

This issue of a general asymmetry in phonosymbolic responses (depending on which pole of the antonymic scale was chosen) does not arise with the two earlier authors whose scales are depicted in Figure 7.1. From Tolman's (1887, 1904) presentation of his tone-color scale, the assumption seems to have been that antonymic endpoints (clusters of attributes) were equivalent in strength. Moreover, in gathering data for his brilliance and size scales, Newman (1933: 62) apparently asked *only* darkness and largeness questions, i.e., his unstated assumption appears to have been that either polar antonym would be equivalent to the other in its power to evoke responses. However, since (at least for the AE data) this appears not to be the case, where necessary, we will focus on salient scale findings as giving a clearer indication of phonosymbolic judgments.

Triggering features. An issue raised at the beginning of Section 7.2 was whether brilliance and energy were in some sense synonymous; a related question is whether the same triggering features were operating in evaluations on both dimensions. A qualified "yes" would be the answer to both questions. Some inter-scale variation must be taken into account, related both to saliency (fluctuation in robustness of response, as discussed above) and to dimension: clear differences in response are seen even on the salient

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

brightness and energy scales (stimulus collapses, index range shrinkage). Yet, especially in the context of the overall high level of consensus, given the similarities in their rankings, together with the difference of these from the size rankings, it is hard to avoid the sense that the brightness/darkness and activity/passivity questions were, essentially focusing subjects on the *same aspects* of the stimulus vowels. This is a hypothesis which will be strengthened or weakened to the extent that we find similarity in brilliance and energy rankings, and difference from those and size rankings, in the data of the other speaker groups.

Thus, the most likely explanation of the rankings is that gradations of one feature were the dominant criteria in generating the four sets. Given the endpoint (rather than middle) ranking of the central vowel <ʌ>, the traditional appeal to anteriority as the dominant triggering feature must be discarded. Instead, the proposed trigger is locus of main concentration of acoustic energy, extended downwards to include <ʌ>'s absence of locus; this accounts plausibly for brightness, and, with different areas of gradational blurring (of the middle range for energy and the lower range for darkness) would also account for rankings on the other three scales. The difference in sharpness of response between energy/brilliance and size and the lessening of the effects of saliency for size can be explained by the lesser dominance of articulatory spaciousness as a triggering feature, and by the likelihood that other conflicting size-triggers, such as perceived duration and perceived intensity, were interfering with its effects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Phonosymbolic profiles of the stimuli.

<a> Non-consensus: 4/18 (22.2%). This stimulus is ranked at an extreme on the two size scales, as largest and least small; however, it is not differentiated from <u> on the size or energy dimensions. Since the collapse of these stimuli is seen across two dimensions whose rankings do not resemble each other's, there is evidence to claim that <a> is generally not well-differentiated from <u>. On the energy and brilliance scales, <a> occupies a middle rank, tending to be judged as somewhat darker and less bright than <i> (the brightest stimulus). Its collapse with <u> on the energy scales suggests that it, like <u>, may have a genuine phonosymbolic middle status—evaluated as neither active nor passive (or having both aspects to it, i.e., features that would trigger both evaluations). In sum, <a> can be said to be a fairly bright vowel, undetermined or neutral with regards to energy, and large.

<i> Non-consensus: 0/18 (0%).This stimulus is evaluated at an extreme on each of the six scales: it is the brightest/least-dark, smallest/least-large, most-active/least-passive. It is always differentiated from its rank neighbor, indicating a decided, phonosymbolically distinct personality. It can be described as very bright, active and small.

<u> Non-consensus: 7/16 (44%). In contrast to <i>, on five out of six scales this mid-ranked stimulus is not differentiated from one of its rank neighbors, <a> or <ʌ>, and on the size scales, from neither. Lack of differentiation from both <a> and <ʌ> occurs across size *and* either energy or brilliance scales (dimensions, as observed above, whose rankings do not resemble each other's). We can therefore claim that <u> does not have a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

very distinct phonosymbolic personality—at least on the semantic dimensions along which it was evaluated in this study. Though occupying a middle rank on all six scales, on brilliance and size, it tends towards the dark/non-bright, large/non-small ends. To sum up, <u> can be said to be rather dark, either undetermined or neutral with regards to energy, and rather large.

<ʌ> Non-consensus: 3/18 (17%) This stimulus occupies a ranking endpoint on four out of the six scales: it is considered the least-bright/darkest and least-active/most-passive. On the darkness scale and both size scales it is not differentiated from <u>. This collapse occurring across two dimensions, brilliance and size, whose rankings do not resemble each other's, gives evidence for a general phonosymbolic association between <ʌ> and <u>. In conclusion, <ʌ> can be described as dark, passive, and undetermined or neutral with regards to size.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

BRILLIANCE, ENERGY AND SIZE IN VOWELS: A CROSS-LINGUISTIC STUDY OF PHONETIC SYMBOLISM

by

STEPHANIE SMOLINSKY

VOLUME II

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2001

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 8

ANALYSIS AND INTERPRETATION OF THE KOREAN DATA

8.1 The phonetic/phonemic background of Korean subjects

In this and the following two chapters, all of which discuss data from the ESL groups, the phonetic background sections will broadly correspond to their AE equivalents in Chapter 7. The ESL participants were expected to have produced interlanguage versions of the stimulus vowels, i.e., pronunciations influenced by those vowels felt as equivalent in their first languages.[1] Thus, Section 8.1.1 will give an overview of the articulatory descriptions of those Korean vowels judged most likely to correspond to the stimulus vowels while Section 8.1.2. will provide their acoustic description. It gives formant values, including values for F_0, and compares them with those of the AE target vowels cited in Section 7.1.2.

8.1.1 Articulatory details

The Korean vowel system can be considered as more elaborate than the five-vowel systems of Japanese[2] or Spanish. Korean vowels may be monophthongal or diphthongal, but it is only the

---

[1] See Chapter 6, Section 6.2. for a discussion of this issue. For discussion of the term *interlanguage*, see Selinker (1972).

[2] This is not counting gemination, which doubles the number of Japanese vowels. However the number of articulatory positions for Japanese vowels is still five.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

monophthongs which need concern us, since it is amongst them that we can find the nearest equivalents to the AE stimulus vowels, both to the tense/diphthongized high vowels [iy] and [uw],[3] and to the non-diphthongized non-high vowels [a] and [ʌ]. Accounts vary, both of the exact number of Korean monophthongs (between 5 and 11),[4] and of their qualities. There are, unfortunately, enough conflicts among the descriptions consulted—especially over the number and nature of the back and central vowels—to make the following survey less definitive than desired.

The point vowels. The vowel [i] is described by Ramstedt (1939: 20) simply as "the high front vowel," as sounding like the *ea* in *treat* by Lukoff (1945: 17), and similarly by Horne and Yun (1950: xv). C. H. Lee (1955: 7) places it slightly lower, in the mid-high position, as sounding like "the vowel in *king*"; Cho (1967: 77) appears to agree with Lee, since he describes [i] as "lax and...the palatal counterpart of [ɪ]." Martin (1992: 24) merely specifies it as high, front and unrounded, while Escardt and Oh (1995) describe it as pronounced like the second vowel in *machine*. None of the authors consulted mentioned either lip spreading or its absence in describing this vowel. Our conclusion must be that Korean [i] is sufficiently close to the AE high front vowel not to have caused problems in producing a similar vowel for <i>, although it may have been articulated slightly lower and perhaps with less lip gesture than its AE equivalent.

---

[3] Even though the English high vowels are diphthongal, their segmental order is VG. Since the Korean diphthongs' order is GV (Martin 1992: 34), they should not have been identified with the English high vowels.

[4] Ramstedt (1939) claims seven, Lukoff (1945) nine, Horne and Yun (1950) eight, C.H. Lee (1955) nine, Cho (1967) eleven, D.Y. Lee (1981) ten, Chung et al. (1988) eight, Yang (1991) ten, Martin (1992) nine, and Escardt and Oh, (1995) seven. Kang (1966) reports dialectal difference, with standard North Korean possessing five, and standard South Korean, seven.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Ramstedt describes Korean [a] as "the normal low-back" vowel (Ramstedt 1939: 20). Lukoff states that it resembles the first vowel in *father* (Lukoff 1945: 15), as do Horne and Yun (1950: xv), C. H. Lee (1955: 7) and Escardt and Oh (1995). However, the reader is reminded that AE [a] is horizontally variable (see Chapter 7, Section 7.1). Lee's (1955) account implies that in the Korean vowel system there is an opposition close to the English [ae]—[a] distinction, but this need not necessarily imply the backing of [a], any more than it does in English. Cho (1967: 80) claims that amongst the Korean monophthongs there is a lax, fronted central [a], but the only back monophthongs are [u] and [o]. He does however describe a monophthong*oid* low back vowel (i.e., an [a] preceded by a very slight labial glide when pronounced at a slow tempo).[5] Martin (1992: 24) appears to agree with Cho, giving a low back rounded [a] on his chart of nine monophthongs (this vowel appears in addition to the rounded [wa] on Martin's chart of diphthongs). In sum, these descriptions suggest that although Korean [a] is consistently low, its anteriority may fluctuate, and it may be accompanied by some tendency to rounding. Thus, there may well have been some variation in how <a> was pronounced by Korean subjects.

Korean, unlike Japanese, has a back rounded [u] sound, but other back, backed-central or central vowels are present in the vowel system, one or more of which may be considered allophonically related to /u/: a high central or back unrounded [ɯ], a mid-high back rounded [ʊ] and even possibly a mid-high back unrounded [ɜ˃]. Thus, Ramstedt (1939: 20) describes a "high-back rounded *u* sound," but adds "it seems to have two shades, or we have two *u* sounds, the one being more back, the other pronounced against the middle of the palate." Similarly, Horne and Yun

[5] *This has some preliminary rounding* "articulated with the tongue position of the vowel [a] and the open lip rounding for the back vowel (as in British *hot*)...of the IPA's secondary cardinal vowels" (Cho 1967: 85).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(1950: xv) imply that the Korean /u/-phoneme is realized as two allophones, one high, one mid-high, describing it as "a sound like the *oo* in *boot*; sometimes like the *oo* in *good*." C. H. Lee, as with his description of Korean [i], identifies [u] as mid-high rather than high, a vowel "similar to the vowel in *full*" (Lee 1955: 7). However, Lukoff describes Korean [u] as resembling the English vowel in *food* (Lukoff 1945: 17), implying it is high back and rounded; Cho (1967: 80) also describes [u] as "back, close, labial, hard [non-palatalized] and tense," and Martin (1992: 24) similarly classifies it as high, back and rounded. Escardt and Oh (1995) also imply that Korean [u] is high back and rounded, since they claim it resembles the AE vowel in *rude*. In sum, a Korean analogue to the AE high back rounded vowel appears to exist, but the presence of other non-front high or mid-high phones, possible variants of /u/, suggests there may have been some variation in the Korean pronunciation of <u>. It also suggests areas of potentential confusion between <u> and <ʌ> for Korean subjects.

The central vowel. Most of the sources consulted report a mid central Korean vowel which corresponds to, and in several cases is actually transcribed as, [ʌ]. Ramstedt (1939: 20) describes a "back vowel...pronounced probably as [in]... *blood, club*," but which is unstable and has merged in many dialects with [a] or with [ɯ]. Lukoff (1945: 17) transcribes a vowel [ø] which has "a sound about like *u* in *cut*"; C.H. Lee (1955: 7) uses the same symbol for a vowel sounding like those in "*up* or *under*." Horne and Yun (1950: xv) similarly refer to a vowel "somewhat like the *u* in *cup*" but use the [ɔ] symbol, perhaps implying a more backed articulation. Kang (1966: 17) describes a central vowel, transcribed as [ʌ], which merges with [o] in the Pyongyang dialect (he does not specify its behavior in the Seoul dialect). Cho (1967: 78-80) describes a mid central

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

velarized phone [ə][6] in which "the tongue root is drawn back and rises in the direction of the soft palate, while the tongue tip is drawn forward...the raising of the post dorsum is...lower than for the vowel [o]." Martin (1992: 24) gives no central monophthongs,[7] but remarks of the mid back vowel that in Seoul speech it has a short allophone transcribable as [ʌ]. Among the "ten Standard Korean [Seoul dialect] vowels investigated," Yang (1996: 248) includes in his acoustic study a phone transcribed as [ʌ].

Further evidence that the AE stimulus vowel <ʌ> could have been perceived by Korean speakers as having an obvious L1 equivalent is given by the fact that the Hangeul character ㅓ may be transcribed by phonetic symbols for the mid central vowels, [ʌ] and [ə]. Ramstedt (1939: 20) does not give an I.P.A. symbol for this character, but identifies it with the central vowel [ɜ] found in the *ir* of *bird*, (in British, non-rhotacized, pronunciation), and, as mentioned above, C.H. Lee (1955: 6) uses the symbol [ø] to transcribe ㅓ . However, in his entry on the "fundamental" Korean vowels, Onishi (1981: 231-232) transcribes ㅓ as [ʌ], classifying it as a back vowel. D.Y. Lee (1981: 185) gives [ə] as the phonetic transcription for ㅓ , a "simple/basic" vowel "spoken in the middle part of Korea, including the speech of Seoul," and Chung et al. (1988a: 226-227) also transcribe ㅓ as [ə]. I asked Mr. Jonathan Bidol, the (balanced) bilingual informant who had checked subjects' translations of the questionnaire adjectives for me, to pronounce the character

[6] The reader is reminded that the [ʌ] vowel in *cup* is often referred to as "stressed schwa" (see Ladefoged, 1982) Moreover, Trager and Smith and other American linguists use [ə] where those following the IPA scheme would use [ʌ].

[7] Cho (1967: 78-80) classifies the high and mid unrounded phonemes, [ɯ] and [ə], as central, while Martin (1992: 24) classifies them as back.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ㅡ both in isolation and in one or two Korean words. My impression was that it was very close to AE [ʌ].

The possibility also exists that for some Korean subjects, their high back unrounded [ɯ][8] may have seemed the closest equivalent to <ʌ>. This could have been a likely choice for speakers of those dialects where [ɯ] tends to merge with a mid central vowel, e.g., as in Chung et al.'s (1988a: 227) observation that "the distributions of... [ə] and [ɯ] uttered by Seoul speakers are separated clearly, but those by Kyungsando speakers overlap heavily." Cho (1967: 78-79) remarks of [ɯ] that it is is the epenthetic vowel for Korean renditions of English words with consonant clusters, as in [sɯteisn] (station); he characterizes it as "the shortest of the vowels" in standard Korean. This is an important phonological/phonotactic characteristic to bear in mind in the course of data interpretations, as notable perceived shortness of a vowel could translate iconically as non-largeness/smallness on the size scales.

As a further check, since several of the descriptions of Korean vowels obtained for this survey are over forty years old, I also asked for vowel descriptions from a professor of contemporary Korean. In a personal communication, Dr. Bruce Ballard, an instructor in Korean at Columbia University, informed me that he teaches the point vowels as follows: [i] has lip spreading, [u] is high back and rounded, and the anteriority of [a] will depend on its phonetic context, i.e. the backness of contiguous consonants. Dr. Ballard did not recognize the high back unrounded vowel

[8] See Lukof (1945: 17), "Korean *u* with the lips spread," similarly, Horne and Yun (1950: xv): "A sound made by keeping the lips spread apart and then pronouncing the Korean u-sound," and C. H. Lee (1955: 7) "Have the tonge well back, place lips in a position intermediate between that for [ʌ] (i.e., relaxed) and [ɪ] (i.e., spread), and make a sound."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

[ɯ] as phonemic (thus implying that it might be a variant of the high back rounded vowel: this resembles Ramstedt's (1939) description, quoted above, of the two types of [u]). However, he also teaches a mid high back unrounded phone, similar to the vowel of *book* except that it should be articulated with a smile (i.e. spreading rather than weak rounding, as in English): this seemed to him to be the nearest equivalent to [ʌ].

To sum up, there is a strong argument to be made that Korean subjects felt that a vowel corresponding to AE [ʌ] was available to them in their first language. But, as has been argued of AE <ʌ>, its articulatory placement may still have varied from subject to subject; in this case, the cause may have been dialectal differences. Thus, Korean <ʌ> could have been articulated as a close, velarized vowel, or with a more open jaw and/or lower tongue position, and it may have been more central or more backed. Such variation suggests that for some Korean speakers, <ʌ> might not have been well differentiated from <u> in height or anteriority.

8.1.2 Acoustic details

In the preceding chapter, acoustic details of AE vowels included the four features considered most likely to have influenced stimulus rankings: formant frequencies, formant intensities (perceived loudness), F_0 values (intrinsic pitch) and intrinsic duration (perceived length). For Korean vowels, however, it has not been possible to locate material on intensity or duration; in consequence, the following discussion treats only formant frequencies and F_0 values. Two fairly contemporaneous modern studies have been located: using data from twenty male speakers, Chung et al. (1988a), give F_1 and F_2 values for the point vowels and for a central vowel

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

transcribed as [ə]. Under half of Chung et al.'s informants spoke the Seoul dialect of Korean.[9] Yang (1996) used ten male informants, all speakers of the Seoul dialect; his study gives values for F_0 through F_3 for the point vowels and for a central vowel transcribed as [ʌ]. Both sets of values are given below, in Tables 8.1 and 8.2. As an aid in the discussion of anteriority, these also include values (in italics) for [o] and for the high unrounded vowel (represented as [ɯ] by Chung et al. and as [ɨ] by Yang). For interlinguistic comparison, the reader will be again referred to Table 7.1 which shows formant values for AE vowels.

As in Chapter 7, in the interests of simplicity, male values only have been used. Chung et al. do not include female informants in their study, and although Yang does, noting some sex- and gender-related differences,[10] statistical testing indicates that these do not appear to have had a phonosymbolic effect on the Korean responses in this experiment. Since the Korean subjects were the only group with an almost even balance of males and females (49 males to 51 females), a Pearson Product-moment Coefficient Correlation Analysis was made of their data. Results gave a correlation of .888037, indicating that there was very little male-female difference between evaluations.

[9] See Chung et al. 1988a. The largest group (ten) were from Kyungsando, while seven came from Seoul.

[10] For details, see Yang (1991: 252, 259).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 8.1: Formant frequencies (Hz) for male Korean speakers

	F_1	F_2
[a]	750	1185
[i]	283	2241
[u]	312	766
[ɯ]	*362*	*1164*
[o]	*436*	*740*
[ə]	515	980

Adapted from Chung et al. (1988a: 228)

Table 8.2: Formant frequencies (Hz) for male Korean speakers

	F_0	F_1	F_2	F_3
[a]	162	738	1372	2573
[i]	172	341	2219	3047
[u]	174	369	981	2565
[ɨ]	*174*	*405*	*1488*	*2497*
[o]	*170*	*453*	*945*	*2674*
[ʌ]	165	608	1121	2683

Adapted from Yang (1996: 251)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The two Korean studies compared. [11] The most striking difference is the consistently higher F_2 values in Yang's study for all five non-front vowels, suggesting an overall more anterior articulation for Yang's vowels and a more backed articulation for Chung et al.'s. However, comparing F_2 values within the two data sets, we see that the horizontal relations between [o] and [u], and between [a] and the high unrounded vowel (Chung et al.'s [ɯ] and Yang's [ɨ]), are similar in both. Very close, lowest F_2 values are found for [u] and [o], indicating backmost articulation, while the F_2 values for the high unrounded vowel are similar to those for [a], suggesting a relatively central articulation for both phones, nearer to each other for Chung et al. than for Yang.[12]

If the F_2 values for the high central vowels are compared with those of the point vowels, they indicate that Chung et al.'s [ɯ] is articulated towards the back (i.e., it is farther in value from [i], closer to [u]), while Yang's [ɨ] is backed-central (its F_2 value being fairly equidistant between that of [i] and [u]). F_2 values also indicate that Chung et al.'s [ɯ], [u] and [ə] are horizontally near to each other; Yang's [ɨ] and [ʌ] are separated, but [u] and [ʌ] are near, and despite more overall anteriority, Yang's [ʌ] is articulated relatively farther back in the system.[13]

---

[11] In this, and all subsequent acoustic comparisons, I have been advised by Dr. Lawrence Raphael (personal communication) that only discrepancies in formant values of around or above 75Hz (for F_1) and around or above 125 Hz (for F_2) should be considered to reflect real differences in the articulation/timbre of the vowels discussed.

[12] Chung et al.'s [a] and [ɯ] have nearly identical F_2 values, while the F_2 for Yang's [ɨ] suggests that it is slightly forward of [a].

[13] The F_2 value for Yang's [ʌ] is closer to that of their [u], while Chung et al's F_2 for [ə] is more or less equidistant in value between those of his central [ɯ/a], and of [u].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A minor but consistent difference, found for all three high vowels, is that Yang's F_1 values are very slightly higher than Chung et al.'s, suggesting, at most, a fractionally lower articulation for Yang's vowels. However, in each case, the disparities are below 75 Hz and thus not likely to reflect substantial variation in vowel height/openness.

The point vowels. Both studies' F_1 values for [a] are similar, suggesting that it is an equivalently low vowel for both sets of speakers. However, since Chung et al.'s value for F_2 is, as noted above, higher than that of Yang's, this indicates either a more central articulation for Yang's [a] than for Chung et al.'s, or that Chung et al's [a] is lip-protruded, in contrast to Yang's. These differences support observations made in the previous section about possible variations in the anteriority and rounding of [a]. In surveying the high vowels [i] and [u], we have already noted that their F_1 values are fairly similar in both studies. F_2 values for [i] are even closer, indicating an agreement on anteriority for this vowel, but, as noted, those for [u] are different: Yang's [u] appears to be articulated more centrally than Chung et al.'s, which is consistent with those variations in Korean [u] discussed in the previous section.

The central vowel. Here, we see disagreement on both first and second formant values, indicating vertical and horizontal articulatory differences. Chung et al.'s [ə] has a lower F_1 value than Yang's [ʌ] and a higher F_2; these indicate that Yang's [ʌ] is a lower (mid low), more central vowel, while Chung et al.'s [ə] is higher, a mid vowel, and farther back..[14] Such observations

[14] Chung et al.'s [ə] has an almost identical F_1 to their mid vowel [e] (Chung et al. 1988a: 228), while Yang's [ɯ] has an F_1 which is close to his mid low [ɛ] (Yang 1999: 251).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

further support the arguments in the previous section that Korean renditions of <ʌ>, may well have, like their AE equivalents, varied in anteriority and height.

Korean formant values from both studies compared with those of AE. For the reader's convenience, Table 7.1 from the previous chapter is reproduced below.

Table 7.1: Formant frequencies (Hz) and intensities (dB, relative to F_0, in parentheses) for male speakers of American English

	F_0	F_1	F_2	F_3
[iy]	136	270	2290	3010
		(-4)	(-24)	(-28)
[a]	124	730	1090	2440
		(-1)	(-5)	(-28)
[uw]	141	300	870	2240
		(-3)	(-19)	(-43)
[ʌ]	130	640	1190	2390
		(-1)	(-10)	(-27)

Adapted from Peterson and Barney (1951: 183)

The point vowels. F_1 values for [a] in the AE and the two Korean studies are almost the same, which suggest similar lowness/jaw opening for this vowel in both languages. However, both of the Korean studies show F_2 values for [a] which are higher than that of AE, indicating that it is articulated with the tongue farther forward than its AE equivalent. The F_1 value for Chung et al.'s [i] is very close to that of Peterson and Barney's, but Yang's is just enough above the AE

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

value to suggest that Korean [i] could vary between having an tongue height equivalent to its American counterpart and having a slightly lower tongue position. This agrees with variations noted above in the articulatory descriptions of this vowel. In both the Korean studies, [i]'s F_2 value is fairly close to Peterson and Barney's, which indicates that it is articulated with equivalent anteriority to AE [i].

Chung et al.'s F_1 for [u] is likewise very close in value to Peterson and Barney's, while Yang's is somewhat above it. This suggests an articulation for Korean [u] varying between a tongue height equivalent to AE [uw]'s and one with a slightly lower tongue position, which again agrees with variations noted in the corresponding articulatory descriptions. Chung et al.'s F_2 value for [u] is fairly similar to that of Peterson and Barney's, while, as observed Yang's is higher; this too indicates variation in articulation between a Korean [u] of equivalent backness to AE [uw] and one with a more forward tongue articulation/less lip protrusion than the AE vowel.

The central vowel. Chung et al.'s F_1 value indicates that their [ə] is articulated higher than AE [ʌ], while Yang's F_1 value for his [ʌ] is close enough to be considered as indicating equivalent height to Peterson and Barney's vowel.[15] AE F_2 values for [ʌ], [a] and [u] indicate that they are positioned centrally, with [a] slightly back of, and [uw] well back of, [ʌ]. Chung et al.'s F_2 values for [ə] and [u] suggest a fairly similar horizontal relation, with [ə] and [a] positioned centrally, and [u] back of [ə], but [ə] somewhat closer to [u]. Yang's F_2 values also suggest an articulation of [u] back of [ʌ], but with [ʌ] even closer to [u]. Thus it appears that the AE tongue position for

[15] Peterson and Barney's (1951: 183) F_1values for their [ʌ] and for the low front [ae] are almost identical, suggesting an articulation between mid-low and low for AE [ʌ].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

[ʌ] is farther forward—and/or that it is less lip-protruded—in relation to [uw] than are Chung et al.'s equivalent Korean vowels. In contrast Yang's central vowel corresponds in anteriority more closely to the AE one. Thus, comparison of formant values suggests that neither Chung et al.'s nor Yang's central vowel is completely equivalent to AE [ʌ].

Locus. In terms of their placements of locii of main concentrations of acoustic energy, Yang's study (the only one to give a range of formant values from F_0 through F_3) indicates that Korean [i] and [u] resemble AE [iy] and [uw]. However, Yang's study also indicates a surprising similarity between [a] and [ʌ] in their absence of locus, and a consequent contrast with AE for those vowels. The reader will recall[16] that in American English, [a] can be distinguished from [ʌ] by the presence and placement of a locus of main concentration of acoustic energy versus its absence. For [a], there is a locus in the mid range, formed by the comparatively closely-valued F_1 and F_2, while there are large differences between the values of F_0 and F_1 and between those of F_2 and F_3. On the other hand, for [ʌ], the values of F_0, F_1 and F_2 are evenly spaced, and thus there is a diffusion of acoustic energy, or an absence of locus. In contrast, Korean has a fairly even spacing of F_0, F_1 and F_2 for both [a] and [ʌ], suggesting a similar diffusion of acoustic energy for both.

Yang's F_0 values compared with those of Peterson and Barney. Of the two acoustic studies examined, only Yang's (1996) gives F_0 values. These are reproduced below, in Table 8.3, while Table 8.4 gives comparable values from the Peterson and Barney study.

[16] For full discussion, see Chapter 7, Section 7.1.2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 8.3: Fundamental frequencies (Hz) for male speakers of Korean

	F_0
[a]	162
[i]	172
[u]	174
[ʌ]	165

Adapted from Yang (1996: 251)

Table 8.4: Fundamental frequencies (Hz) for male speakers of American English

	F_0
[a]	124
[iy]	136
[uw]	141
[ʌ]	130

Adapted from Peterson and Barney (1951: 183)

As noted in Chapter 7, the AE values show a stepwise spacing, each vowel evenly separated from its neighbor by small intervals (of 5 or 6 Hz), yielding the order (lowest to highest) [a], [ʌ], [iy], [uw]. The Korean F_0 values yield the same order, but the spacing between them is different. Those intervals between the two high and the two non-high vowels are very small (respectively 2 and 3 Hz) while there is a larger interval (7 Hz) between the two pairs of vowels. This suggests that if the Korean ranking data were to show a similar configuration (<a/ʌ> and <i/u> collapses, with an interval between the collapsed pairs), then evidence might be found in favor of the phonosymbolic effect of inherent pitch.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

8.2 Analysis of the Korean data

Figure 8.1 below presents in summary form the stimulus rankings on six semantic scales, derived from the responses of the Korean speakers. It corresponds exactly to Figure 7.2, summarizing AE responses, in Chapter 7. Thus, it is based on ranking indices, is accurate in scale, and rankings are again integrated with consensus data, with only the more unusual cases of non-differentiation indicated.

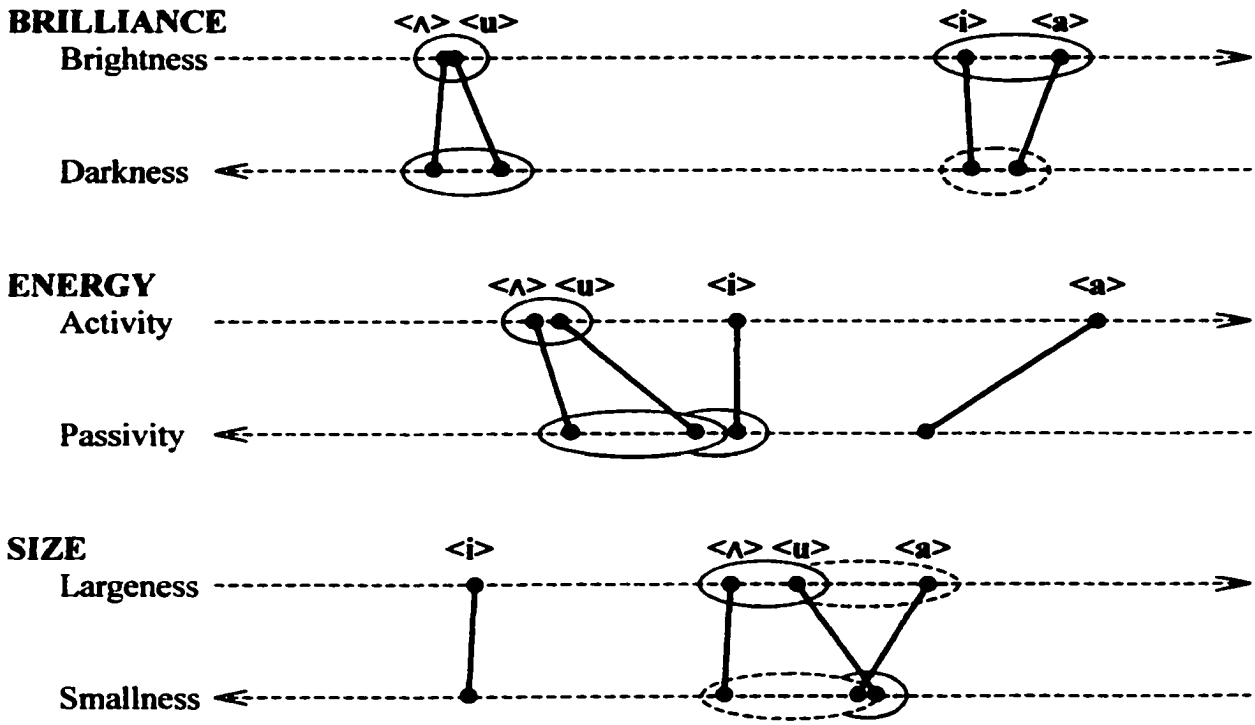


Figure 8.1: Data patterns (Korean speakers) Ranking index values and consensus data for salient and non-salient sub-scales of three semantic dimensions. Ellipses surround stimulus pairs not differing by statistical test, $p>.05$; dashed outlines indicate instances of trend, $.10>p>.05$.

Inspection of Figure 8.1, as with the AE data, reveals a broad similarity between the sets of stimulus rankings obtained on the brilliance and energy dimensions; all of the relevant scales

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

pattern together. However, the rank order of stimuli is not exactly the same as that found in the AE responses: strikingly, Korean subjects judged <a> (rather than <i>) as brightest/least dark and most active/least passive. On the other hand, <ʌ> at the opposite endpoint is judged by both Korean and AE speakers as the least bright/darkest and least active/most passive. Also, as found in the AE data, the size dimension displays a different stimulus ranking from the other two, and we note that the Korean size rankings are in complete agreement with those of the American English speakers. Although <a> is again ranked at an endpoint, as largest/least small stimulus, the opposite endpoint is occupied by <i>, while <ʌ> falls in a middle position, in contrast to its endpoint position on the brilliance and energy scales.

The Korean ranking similarities between the brilliance and energy scales are the second interdimensional agreement of this type to be found; the indication is thus strengthened that some semantic metadimension, i.e., including brilliance and energy as subdimensions, was being accessed by both Korean and AE subjects. Once again, the Korean data provide a reasonable basis for categorization of the six scales into a brilliance/energy grouping of four and a size grouping of two; this possibility will be further explored in what follows.

Symmetricality of rankings for salient and nonsalient scales is found, with one minor exception.[17] The non-salient scales have generally narrower spacing than the salient, but spacing/range disparity is minimal for the brilliance scales, while energy shows a very striking one, and size, a notable but lesser disparity. Since these subjects were not native English speakers, the question now arises whether spacing/range disparities could have been the result of failing to translate

[17] The exception arises with the stimuli <a> and <u> on the size scales.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

accurately any of the questionnaire's six antonymic terms. Fifty of the one hundred Korean subjects responded to a section (added to the questionnaire when the experiment was already well under way)[18] requiring native language translations of the terms. These were then back-translated by a Korean speaker bilingual in English and Korean. Results indicated that the English antonym pairs were correctly translated into their equivalent Korean semantic opposites 84% of the time for small/large and dark/bright and 94% of the time for active/passive, differences in accuracy which do not correspond to those outlined above in spacing/range disparity on the three dimensions. Therefore, in the Korean data, asymmetries cannot be considered as due to mistranslation of the terms used on the questionnaire.

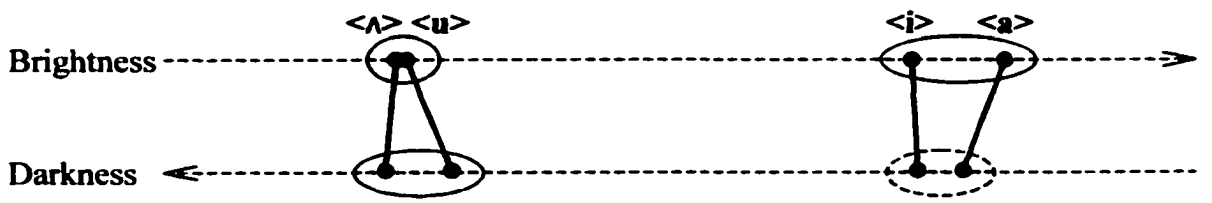
A third and final point is, once again, the importance of stimulus contrasts arising from the pairings of stimuli in the test items. As in the AE case, evaluation of a stimulus seems to have been affected not only by the scale/dimension it was judged on, but also by which other stimulus it was being compared with. In addition, the Korean data show evidence of an interaction between both of these factors. Thus, Figure 8.1 indicates that three stimulus pairs, <ʌ/a>, <ʌ/i> and <u/i>, yielded well-differentiated evaluations, with consensus across five or all six of the scales. An interaction between dimension and pairing is shown by two pairs: <a/u>, with consensus for all except the size scales, and <i/a> with consensus for all except the brilliance scales. Finally, the pair <ʌ/u> is overall poorly differentiated, yielding what could be described as a *cluster* of non-consensus, on five out of the six scales.

[18] For details, see Chapter 6, Section 6.2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The data will now be treated in detail, following the scheme laid down in Chapter 7. Subsection 8.2.1 will discuss brilliance, 8.2.2 energy, and 8.2.3, size, while 8.2.4 summarizes findings and presents a semantic profile of each of the stimulus vowels as evaluated by the Korean subjects. Issues outlined above, arising from initial observations of the data patterns overall, will be explored as they emerge in analyses and discussions which examine the findings more closely.

8.2.1 Brilliance dimension



Brightness

	<u>	<a>	<i>
<ʌ>	48 / 51	95 / 5	88 / 11
<u>		93 / 6	85 / 15
<a>			44 / 56

Darkness

	<u>	<a>	<i>
<ʌ>	48 / 51	7 / 93	9 / 91
<u>		20 / 80	12 / 88
<a>			59 / 40

Figure 8.2, Brilliance dimension data arrays (Korean speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel*: Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing signficicantly, p>.05; dashed outlines indicate instances of trend, .10>p>.05.

Following the form of the previous chapter, Figure 8.2 above extracts the brilliance dimension data previously presented as part of Figure 8.1's overall summary of the data for Korean speakers

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and combines it with the frequency matrices from which it derives, permitting a detailed consideration of subjects' evaluations of the stimuli on brightness and darkness.

General remarks on the scales. The upper panel of Figure 8.2 shows that on both scales, the rankings are collapsed into two 'steps,' with <ʌ/u> judged undifferentiatedly as dark/nonbright stimuli and <a/i> as bright/nondark ones. Also, both scales' index value ranges are similar: while there is a small difference in favor of the salient scale, the shrinkage in range is small (.029 points). Finally, the lower panel of Figure 8.2 indicates a tendency to robustness of response on the non-salient scale. Although there is a modest difference between responses for <ʌ/a> and a greater one for <a/u> in favor of the salient scale, for <ʌ/u> there is no difference, while for <ʌ/i>, <u/i> and <a/i>, it runs in the opposite direction.

The Korean brilliance data thus contrast strikingly in three ways with those of the AE subjects. Firstly, the spacing configurations are different, since the AE results, especially on the salient scale, show even spacing between the four stimuli, i.e. four steps, as compared to the Korean scales' two. Secondly, there is a much larger difference in values between the index ranges of the two AE scales (.096 points), while thirdly, in the AE data, it is the *salient* brilliance scale which shows a tendency to robustness of response. Yet in one aspect, the Korean and AE brilliance data show similarity. As with the AE subjects, these two scales evoked the widest index ranges of the six presented to this group, suggesting that both sets of subjects found questions on this dimension the easiest to answer.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Interpretation of stimulus rankings. The most remarkable finding here is the collapse of <a/i>, ordered as brightest/least dark stimuli, and the ranking of <a> as the brighter/less dark of the two. This ordering for <a> conflicts with the findings of Tolman (1887, 1904; see also Chapters 2 and 7)[19] and Newman (1933), and the views of all of who consider [i] to be the brightest vowel (see, e.g., Jespersen, 1933, Grammont, 1933, Chastaing ,1962, et al.). It also shows disagreement with the AE group's ranking of <i> as brightest/least dark stimulus. Yet there are grounds for finding the Korean rankings less anomalous than they might first appear. Even though [a] does not have the concentration of high frequency acoustic energy that [iy] or [i] does in English, Korean, Japanese or Spanish,[20] it can be argued to possesses other characteristics capable of triggering a brightness judgment: high acoustic amplitude and its articulatory correlate, wide jaw opening. Thus, the collapse of <a> with <i> for the Korean subjects can be explained as a way of resolving a contest between two brightness triggers. For these speakers, the amplitude feature (loudness/openness = brightness) competes with the frequency feature (high locus/lip spreading = brightness), and the <a/i> non-differentiation represents a compromise. In contrast, the AE ranking of <i> as brightest and <a> as second-brightest can be seen as ranking the two brightness-connoting features in an order of linguistically- or culturally-based importance.

In addition, another factor may have served to associate <a> with <i> in terms of brilliance, and to differentiate these stimulu from <ʌ> and <u>. In investigating the acoustic aspects of Korean [a] and [ʌ], both studies consulted (Chung et al., 1988a and Yang, 1996) report higher F_2 values

[19] However, Tolman evaluated [iy] as fourth-brightest/happiest/smallest/quickest vowel (see Chapter 7, Figure 7.1). This seems to agree with its evaluation by the Korean subjects as tending to be bright, but not the brightest stimulus.

[20] *For details, see the relevant acoustic background sections* (Sections 7.1.2, 8.1.2, 9.1.2 and 10.1.2.)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for [a] than its AE equivalent (Peterson and Barney, 1952), suggesting that it is articulated farther forward than AE [a].[21] On the other hand, the relatively low F_2 of Korean [ʌ] reported in Chung et al.'s (1988a) study supports the claim that for some Korean subjects, it may well have been a back, rather than central, vowel. Consequently, sense of a front-back division between the stimuli[22] may have suggested pairings in a brilliance polarity: the non-back <a> was considered as a front vowel along with <i>, and therefore bright, while <ʌ> and <u> were both considered back vowels, and therefore dark.

In further accounting for the <ʌ/u> collapse, it has been hypothesized (see above, Section 8.1.2), that for speakers of some Korean dialects these vowels may not be phonemically distinct, or that the presence of [ɯ] as an allophone may overlap both. In addition, at least one authority (Cho, 1967) describes [ʌ] as velarized, which implies both height and backing, i.e., characteristics which would associate it with [u]. We note that AE speakers were also liable to collapse <ʌ> and <u>, and indeed did so on the darkness scale; in their case too, a tendency to back and velarize <ʌ> was the suggested cause. However, the locus feature-equivalence argument (i.e., the equivalence between [u]'s low locus of concentration of acoustic energy and [ʌ]'s absence of such locus, suggested as responsible for the AE <ʌ/u> darkness collapse) will not explain the Korean <ʌ/u> collapse. This is because for Korean speakers, [ʌ]'s absence of locus appears to

[21] For a full dicussion of these issues, see Section 8.1.2, above.

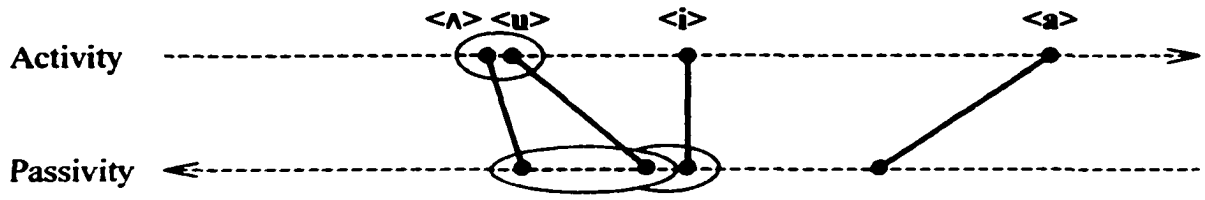
[22] Macdermott (1940) maintains that anteriority is *the* most important phonosymbolic distinction in English. For general discussion of the phonosymbolic association between anteriority and brightness, see also the previous chapter (Section 7.2.).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

have been shared by [a];[23] if locus were a critical brilliance trigger, we would predict a collapse of both <ʌ> and <a> with <u>, instead of the actual opposition between <ʌ/u> and <a>.

8.2.2 Energy dimension

Figure 8.3. presents the data arising from judgments on the energy scales by Korean speakers.



Activity

	<u>	<a>	<i>
<ʌ>	50 48	89 9	65 35
<u>		88 12	63 37
<a>			23 77

Passivity

	<u>	<a>	<i>
<ʌ>	44 54	25 74	33 66
<u>		37 63	47 53
<a>			68 31

Figure 8.3: Energy dimension data arrays (Korean speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel*: Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing signficicantly, p>.05.

General remarks on the scales. While symmetricality of rankings is maintained, we see here several striking differences between the salient and nonsalient scales. Firstly, the configurations

[23] See Yang 1996 and the discussion in Section 8.1.2 above.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in the upper panel of Figure 8.3 indicate that both scales have three steps (cf. the two-step configurations of the brilliance scales), but that passivity alone places <u> indeterminately between the lowest and middle step, i.e. that the collapsed pairs <ʌ/u> and <u/i> constitute a *chain* whose endpoints, <ʌ> and <i>, are differentiated. Secondly, there is a much greater spread of index range values on the salient than on the non-salient scale (of .852—.311 versus .655—.312 index points, with the reduction occuring in the upper values). Consequently, there is a large range shrinkage (of .198 index points) between the two scales. Finally, the lower panel of Figure 8.3 indicates that there is a decrease in robustness of response on the non-salient scale for the majority (4/6) of the six stimulus pairs; exceptions are <ʌ/i>, showing little differences in its scores, and <ʌ/u>, the robustness of whose scores decrease on the salient scale.

Energy compared to brilliance. A feature consistent over both scales on both dimensions is the collapse of the pair <ʌ/u>, but there are also several striking intermdimensional differences. Firstly, in contrast with their collapse on brilliance, <a> and <i> are now ranked on steps independent of each other: <a> remains at the endpoint, while <i> travels downward, from the upper to the middle range. Secondly, while the brilliance scales are similar in spacing configurations, the energy dimension shows large differences between the salient and non-salient configurations, with passivity tending to pull three stimuli together in the above-mentioned chaining of <ʌ/u> and <u/i>. Thus, on the passivity scale, we see a unique opposition of the triad, <ʌ>, <u> and <i>, to non-passive <a>. The chaining produces two specific contrasts with the brilliance rankings: one is in <i>'s collapse with <u> and a second is <u>'s intermediate placement stepwise between <ʌ> and <i>. Third and finally, we see a striking difference in salient vs. non-salient range disparities for the two dimensions: both are in the same direction,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with shrinkage on the non-salient scale, but that for energy is far larger than that for brilliance. To aid the reader, Table 8.5 below makes this clear.

Table 8.5: Ranking index ranges and range discrepancies, for salient versus non-salient sub-scales of two semantic dimensions

	Salient	Non-Salient	Discrepancy
Brilliance	.591	.562	.029
Energy	.541	.343	.198

Interpretation of stimulus rankings. Even more plausibly than with brilliance, <a>'s ranking as most-active/least-passive can be explained as due to its acoustic intensity/openness: there is an obvious iconic association to be made between loudness and activeness. Professor William Stewart (personal communication) has also remarked that maximum mouth-opening for speakers of Asian languages may be associated with *assertiveness*, with emphasizing the self; and a connection can also be seen between assertiveness and activeness.[24]

The position of <i> now has to be accounted for, namely its movement away from the collapse with <a> at the endpoints of the brilliance scales. It can be argued that while the <a/i> brilliance collapse is due to a feature equivalence (high locus/lip spreading = brightness, and acoustic amplitude/openness = brightness) on the energy scales, these features no longer have equal weight. Yet we have previously noted, in Section 8.2, the striking similarities in interdimensional agreements between brilliance and energy rankings, found in both the English and Korean data

[24] Professor Stewart has further remarked that the symbolization of assertiveness by mouth openness is probably a cultural areal feature. Thus, we might predict similar rankings for <a> in the Japanese brilliance and energy data.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sets, which suggest that both subject groups may have felt that on some level brilliance and energy judgments were similar. How can these observations be reconciled?

Let us hypothesize that on both dimensions, evaluation questions did indeed focus subjects on those features which would connote a common quality—a metadimension—which we can provisionally name "vitality." Yet even so, the actual pairs of terms *bright-active* and *dark-passive* are not synonymous; thus, e.g. *bright* might still tend to highlight one phonetic feature and *active* another. In other words, subjects could have responded both to a *metadimensional relation between brilliance and energy* (leading them to search for common, or equivalent, features connoting vitality), and the *particularity of each of the four terms involved* (leading them to look for features specifically connoting, e.g., brightness or activeness).

If this hypothesis is correct, then, applied to the Korean case, we can further develop the argument relating to <i>'s position. The locus features of <i> may have continued to move it towards the *more-vital* (active/non-passive) ends of the energy scales, but they do not connote activeness as effectively as they do brightness; possibly, their iconic connection is not as clear.[25] On the other hand, if <a>'s high amplitude/openness do, as proposed, strongly connote activeness, then <i>'s lower amplitude/closeness (i.e., its lesser 'assertiveness,' compared to <a>) can be argued as connoting its opposite, moving it towards the non-active/passive ends of the scales. In other words, we see a feature conflict for <i>, with amplitude features reducing the effect of locus features. Thus, on the energy dimension, amplitude/openness dominates high

[25] This is meat for another dissertation, on sound symbolism rather than phonetic symbolism. Why is it so generally acknowledged that high frequency acoustic energy connotes brightness and low frequency energy, darkness?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

locus/lip spreading, and <i> is not only relegated to second place, but travels far enough downwards on the passivity scale to collapse with <u>, a unique occurrence for this pair in the Korean data. The <u/i> collapse, and more generally, the chaining of <ʌ>, <u> and <i> on the passivity scale, can now be explained as due to a shared lack of amplitude/openness, found in all three stimulus vowels and opposing them to <a>. This argument assumes that <ʌ> is classified by Koreans as either high or non-low (in contrast with the low <a>).[26]

The indeterminacy of <u> can be seen as the result of features shared with each of the endpoint stimuli of the chain. With <i>, the most likely candidates are low acoustic amplitude/closeness; with <ʌ> (see Section 8.2.1 above), they are backness and velarization, plus a possible allophonic relation. On the other hand, it is more difficult to explain the differentiation of <i> from <ʌ>. If it were due to <ʌ>'s non-anteriority, then <i> should also have been differentiated from the non-anterior <u>, while if it were due to differences in height/closeness, then the (presumeably) mid-high/close <ʌ> would be differentiated not only from <i> but from the equally high/close <u>. No general explanation will fit all the facts here; instead, we must look at the individual stimulus pair[27] for an explanation of those features which may have contrasted them. In the case of <i> and <ʌ>, potential contrasting features are, firstly, <ʌ>'s lack of locus/neutral lip gesture versus <i>'s high locus/lip spreading (proposed as a general vitality trigger) and secondly, <ʌ>'s

[26] The reader will recall that in the articulatory account of Korean vowels (Section 8.1.1), it was proposed that some of the subjects may have associated the stimulus vowel <ʌ> with their own high unrounded [ɯ]. Also, Chung et al.'s (1988a) F_1 value for [ʌ] indicate that it is a non-low vowel.

[27] The reader is reminded (see Chapter 7, Section 7.2.3, for full discussion) that though the rankings show all four stimuli on a single scale, this does not necessarily mean a unification of evaluating criteria. For subjects, judgments were made as isolated acts, *on one stimulus pair at a time, without reference to the other stimuli.* Therefore, it was possible for one pair to be evaluated on the basis of one or more features, while another pair could be evaluated on a different feature basis, a third according to different criteria, and so on.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

epenthesizability (i.e. its phonological/phonotactic shortness,[28] possibly connoting neutrality or insignificance) versus <i>'s full vowel status. We have already noted that for this group <ʌ/i> are well differentiated (on 6/6 scales), which suggests that the features which distinguish them are so notable as to seem relevant in all semantic contexts.

8.2.3 Size dimension

Figure 8.4. presents the data arising from judgments on the size scales by Korean speakers.

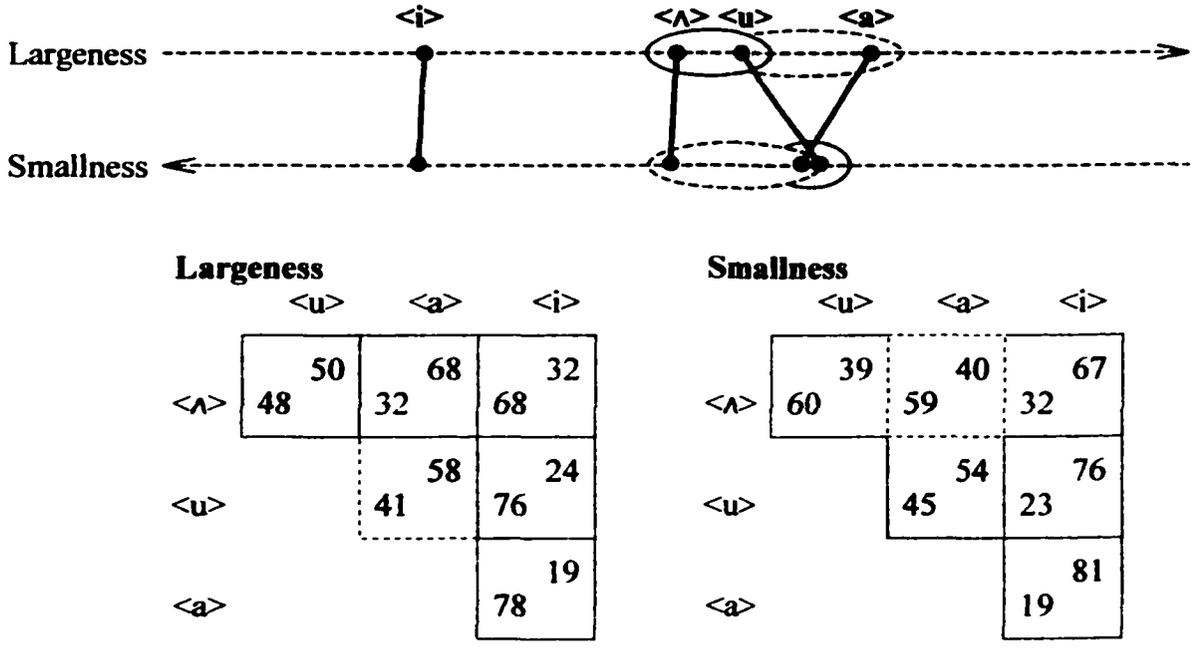


Figure 8.4: Size dimension data arrays (Korean speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel*: Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing signficicantly, p>.05; dashed outlines indicate instances of trend, .10>p>.05.

[28] This is predicated on Korean subjects' equation of the stimulus vowel <ʌ> with their own high unrounded [ɯ]. See remarks above in Section 8.1.1, also Cho (1967: 78-79).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

General remarks on the scales. The upper panel of Figure 8.4 indicates that both scales have the same three-step spacing configuration: one stimulus, <i>, is evaluated as least-large/small, while the three remaining larger/non-small stimuli form chains. On the salient scale, the chain is composed of the collapsed pairs <ʌ/u> and <u/a>, with <u> indeterminate between largest and second-largest step, and on the non-salient scale, the chain is formed by <ʌ/a> and <a/u>, with <a> indeterminate between least-small and second-least-small steps. Thus, while the collapse of <a/u> is consistent for both scales, the rankings for these stimuli are different. We see a restricted spread of values on both scales; their index ranges are fairly narrow, and there is only a small difference between them (a discrepancy of .044 points). The lower panel of 8.4 suggests that generalizations cannot be made about salient-nonsalient differences in robustness of response. For two stimulus pairs, <ʌ/a>, <u/a>, response is in favor of the salient scale; for two more pairs, <ʌ/u> and <a/i>, it is in favor of the non-salient scale, while there is no difference in the repsonses for <ʌ/i> or <u/i>.

Size compared with brilliance and energy. In the type of configurations found on them (chaining versus one independently-ranked stimulus), the size scales resemble the passivity scale, which also has the narrowest range of the six and consequently the greatest amount of compression. However, the ordering and composition of the size scale chains are different from those of the passivity scale, and the ranking asymmetry found on them (of <u> and <a>) is the only case in this data set. Notable is the independent endpoint ranking of <i> on both size scales; elsewhwere, <i> is independently ranked only on the activity scale, and in the middle range. Finally, the smallness scale is unique in showing a differentiation between <ʌ> and <u>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Range data suggest that the Korean scales can be divided into two sets of three, with brightness, darkness and activity obtaining a wider spread of values, while largeness, smallness and passivity show a narrower spread. As an aid in comparison, Table 8.6 below gives the index value ranges for the six Korean scales.

Table 8.6: Ranking index ranges and range discrepancies, for salient versus non-salient sub-scales of three semantic dimensions

	Salient	Non-Salient	Discrepancy
Brilliance	.591	.562	.029
Energy	.541	.343	.198
Size	.436	.392	.044

Differences in range widths suggest that the Koreans, like the AE subjects, found brilliance an easy dimension on which to make judgments, and size more difficult. For this group, saliency had relatively little effect on either size or brilliance, though the non-salient scales appear to have been somewhat more problematic (which agrees in general tendency with the AE data). In contrast, for energy, we see an interaction between dimension and saliency. Judged by its range, the salient energy scale appears to have been almost as accessible for judgments as the non-salient brilliance scale (this is similar to the AE results). However, the narrowness of the non-salient energy range suggests that it was even less easy to judge stimuli on than either of the size scales (which is different from the AE findings, although AE energy also shows the highest salient versus non-salient discrepancy in ranges).

Interpretation of stimulus rankings. Given that the AE and Korean size scale configurations are very similar, these suggest a similar explanation: that once again there may not have been a dominant size trigger, and that more than one feature of the vowels, or interactions between the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

features, may have triggered size judgments. In surveying the five potential size-connoting features, we have noted that for two, vowel duration and amplitude, studies yielding values could not be located for Korean. However, the ordering of certain stimuli may still be argued for, if speculatively ranked according to these features. Table 8.7 below gives, therefore, tentative duration and intensity rankings for certain of the stimulus vowels, together with actual, value-based rankings for the other three possible size triggers; these are compared with the rankings obtained via the data from this experiment. We speculate, following Cho's remarks about the nature of [ɯ], that for speakers identifying <ʌ> with it, this would have seemed the shortest vowel, while <a>, having the widest mouth opening, was likely to have been the longest; thus, they are assigned the respective values of 1 and 4. As noted above, (see Section 8.1.2) Yang's (1966) formant values suggest grounds for a possible lack of differentiation between <ʌ> and <a>, in that they share a lack of locus of main concentration of acoustic energy: this is reflected in the table by their shared (least-ranking) value of 1. Similarly, since Yang's F_0 values suggest paired collapses, i.e., (<ʌ/a> on the first rank step and <i/u> on the second), the first pair are assigned a value of 1 and the second a value of 2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 8.7: *Observed stimulus ranking for size dimension (order being specified as least size = 1 through most size = 4), compared with rankings predicted by potential size-triggering features*

Observed Rankings	**1**	**2**	**3**	**4**
Largeness	<i>	<ʌ>	<u>	<a>
Smallness	<i>	<ʌ>	<a>	<u>
Articulatory spaciousness	1	2	3 *(or 4)*	4 *(or 3)*
Perceived duration	?	1	?	4
Perceived intensity	?	?	?	4
Intrinsic pitch	2	1	2	1
Energy concentration locus	2	1	3	1

Since the Korean and AE size scale configurations are the same for largeness and differ only in the ranking of <a>on smallness, many of the individual arguments put forward for the AE size results are relevant here. We note that as least-large/smallest stimulus, <i> is evaluated by Korean subjects exactly as might have been predicted by earlier, Eurocentric writers on the subject.[29] The opposition of <i> to the chains <ʌ>-<u>-<a> and <ʌ>-<a>-<u> suggests once again that the traditionally-cited feature of articulatory spaciousness may have been dominant, while the <u/a> collapse on both scales suggests a (sub) feature equivalence between horizontal and vertical spaciousness, both connoting largeness. The differentiation of <i> from its rank neighbor <ʌ> can be ascribed to contrasts in one or both of two features. Firstly, in terms of their relative (horizontal) spaciousness, the non-anteriority of <ʌ> (whether it was pronounced with a

[29] See, e.g., Jespersen (1922), Sapir (1927, 1929) and Newman (1933).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

more central or more back tongue position) may have contrasted with <i>'s anteriority, and secondly <i>'s high locus of main concentration of acoustic energy may have differentiated it from <ʌ>'s absence of locus.

Smallness: the unique differentiation between <ʌ> and <u> and unique collapse of <ʌ/a>. We have noted (see Section 8.2) that there is a cluster of non-consensus for the pair <ʌ> and <u> (5/6 scales), and that this occurs nowhere else in the Korean data. Shared features suggested as triggering the <ʌ/u> collapses include backness and velarization/height, and for speakers of some Korean dialects, a possible allophonic relation between their versions of <ʌ> and <u>.[30] Certainly, shared backness will fit with an argument that the stimuli are size-ranked on the basis of articulatory spaciousness, but this will not explain the striking differentiation[31] between <ʌ> and <u> on the smallness scale. However, if some Korean subjects associated <ʌ> with their own central/back unrounded [ɯ], the term *small* may have foregrounded language-specific phonetic/phonotactic features of [ɯ][32] which focused those subjects on its notably short perceived duration, shortness connoting smallness.[33] This distinction may have moved <ʌ> away from <u> and the other non-small stimulus, <a>.

---

[30] However, we note that the AE data show a similar but less widespread pattern of <ʌ/u> collapses (3/6 scales), suggesting that failure to differentiate between these stimuli may not have been a language-specific phenomenon.

[31] Although more pronounced in the AE than the Korean data, we are beginning to see a general tendency for more differentiations on salient scales and more collapses on non-salient scales. Unusual differentiations like this might be considered to carry *especial weight* when they occurr on a non-salient scale, and the same could be said of unusual collapses on a salient scale.

[32] For details, see Cho (1967: 79).

[33] For the association between short perceived duration and phonosymbolic attribuations of smallness, see the discussion (in Chapter 2) of the findings of Sapir (1929) and Newman (1933), and the discussion of the AE size

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Several shared features may have contributed to the <ʌ/a> collapse. Their F_0 values are close (both lower, arguable as connoting non-smallness). Also, both lack a locus of main concentration of acoustic energy; this could be seen as maximally opposing them to <i>, which has the highest locus and is judged the smallest stimulus. In addition, Yang's (1996) formant values sugggest that Korean [ʌ] is a fairly open vowel.[34] If horizontal articulatory magnitude were especially highlighted by the smallness measure, then what potentially connoted non-smallness about <ʌ>, its non-smallness, may for some speakers have made it seem more like <a>, and less like <u>.

8.2.4 The relevance of Martin (1962) to this study, concluding remarks on the Korean data and phonosymbolic profiles of the stimulus vowels according to the Korean subjects.

Martin's heavy/light phonosymbolism, compared with the present findings. Martin (1962) appears to be the only English-speaking authority to have written on Korean phonetic symbolism; his views cast an interesting sidelight on the findings of the present experiment. His system is based on observations of minimal or near-minimal pairs of expressives[35] which vary semantically as a segment or segments in the word are changed, recalling Sapir's (1927) original *mil* and *mal* examples.[36] Korean monophthongs are classified into four phonetic series: front rounded, front

---

results in Chapter 7 (Section 7.2.3).

[34] See Yang (1996: 251), and Section 8.2.2 above.

[35] Expressives, also referred to as *mimetic words* (Hamano 1986, 1994), *descriptives*, *impressifs*, *ideophones*, etcetera: for a list of terms, see Diffloth (1994: 108), are a syntactic class of words, frequently compared to adverbs, found in certain African and Asian languages, including Japanese and Korean. These words use systematic phonetic symbolism to perform such functions as modifying a verb of general action like "go" to produce specialized forms, like "waddle," "lumber" and "flit." This systematization means that speakers can coin and instantly understand new expressives, and has allowed linguists, amongst them Martin (1962), Hamano (1986, 1994) and Diffloth (1994), to study phonetic symbolism in the languages in which such words occur.

[36] Martin (1962: 183) terms these "word isotopes. Some Korean words have several shapes with different vowels. For the meaning 'with a moan or a groan' there are four shapes: *kkayng*, *kking*, *kkong* and *kkwung*. *Kkayng* and *kkong* are used for light quiet moaning and *kking* and *kkwung* for heavy, loud groaning... the differences that we have

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

unrounded, back rounded and back unrounded. They are also cross-classified phonosymbolically as "heavy" or "light," according to their degree of openness, the lowest, most open member of each of the series being the light vowel,[37] while the vowel(s) above is, or are, heavy. Other features, such as anteriority or rounding do not count, except in the case of the high unrounded vowels, [i] and [ɯ], which may either be heavy *or neutral*. This further suggests that <i> (and also <ʌ>, if felt by Korean subjects to correspond to their [ɯ]) may have been evaluated in a way indicating semantic neutrality.

Martin's grouping of vowels is as follows:

LIGHT	HEAVY
[ø] [æ] [a] [o]	[i]* [e] [ɯ]* [ɜ] [u][38]

He speculates that the heavy-light division may relate to a basic Taoistic worldview, but with a reversal of the expected associations. [39] Martin lists a range of positive and negative meanings[40]

---

managed to pin down I refer to in terms of 'word isotopes': a given word may have a LIGHT isotope and a HEAVY isotope. Certain vowels ([a], [æ], [o],[ø]) are typical for light isotopes, while others ([ɜ], [e], [u]) are typical for the heavy isotopes. The vowels [ɯ] and [i] are either heavy or neutral."

[37]This suggests that the light vowels may be the marked ones, while the heavy vowels are in the "elsewhere" category. This analysis is further supported by Martin's claim that [i] and [ɯ] can be heavy *or neutral* while he reports no light/neutral vowels and also from further commentary on the restricted distribution of light vowels in expressives: "The isotopic difference of meaning is not always strongly felt, and sometimes one isotope is felt to be 'neutral' or 'basic'—frequently the one with the 'heavy' vowel unless the underlying meaning is strongly suggestive of the 'light' vowel." (Martin 1962: 183).

[38] Asterisks indicate possible neutrality.

[39] Thus the Chinese philosophic system associates the concept of *yin* with darkness and femaleness and that of *yang* with brightness and maleness, while the Korean phonosymbolic scheme associates *um* (*yin*) with heaviness, substantiality and maleness and *yang* (*yang*) with lightness, insubstantiality and femaleness (Martin 1962: 184). "We are tempted to link up the vowel symbolism with the Chinese concept of *thayuk* 'the great ultimate' which divides into *um* 'the dark side' and *yang* 'the bright side.' But in Chinese thinking, the male element is 'bright' and the female 'dark,' so the Korean concept seems askew here, just as in its choice of lower vowels for the light isotopes and higher vowels for the heavy isotopes, Korean runs counter to the notions of Edward Sapir, Roger Brown and many others, with respect to the non-culturally determined 'naturalness' of phonetic symbolism. Moreover, the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

associated with the original heavy-light vowel qualities, and a tendency to gender-differentiated use.[41]

Given the four stimulus vowels used in this experiment, Martin's heavy-light scheme should have translated into an opposition between <a>, the only stimulus with an open (light) vowel, and <i>*, <u> and <ʌ>(*), the three with close or half-close vowels (heavy, or heavy/neutral, as indicated by asterisks). Also, for <i> (and possibly for <ʌ>),[42] there may have been a tendency to act as a semantic neutral on the scales, so that "neither light nor heavy" could translate as, e.g., "neither bright nor dark." Neutrality might be indicated either by consistent middle ranking of a stimulus[43] or by its tendency to collapse with more than one[44] of the three possible partners, or by both.

The configuration most closely agreeing with Martin's scheme is found on the passivity scale, with <a> clearly opposed to the chain formed by the other three stimuli. If lightness can be seen

---

Chinese idea centers on brightness and energy or vitality, while the Korean seems to focus more on substance" (Martin 1962: 184).

[40] These recall Fónagy's (1979) description of the radiating extensions of core metaphoric qualities (see Appendix 6).

[41] Light vowels are used by or of women and children with positive meanings such as "petite," or "delicate" and negative ones such as "dinky," "petty," "flighty," or "silly." Heavy vowels are used by or of men positively to mean "ample" or "weighty" and negatively to mean "dark," "gloomy," "bulky" or "inaccessible." (Martin 1962: 183-184).

[42] The reader will recall that this would be the case if Korean subjects had equated <ʌ> with [ɯ].

[43] But since there are only six scales, a very high majority, i.e., consistent rankings on at least five out of the six would be needed to constitute solid evidence for or against the neutrality of a vowel. In the case of differentiation or its absence, since there are eighteen possibilities (three stimulus pairs multiplied by six scales) a somewhat lower majority might still give sufficient evidence for or against neutrality.

[44] Lack of differentiation from a single stimulus pair partner would merely indicate that subjects may have had a tendency (phonetically or phonologically) to associate one stimulus vowel with a specific other stimulus vowel. Lack of differentiation from two or more partners could indicate a more general lack of fixed semantic identity for a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

as connoting activity, and heaviness, passivity, then the rankings can be explained: <a> with its "dinky, insubstantial" vowel is judged as non-passive, while the <ʌ>-<u>-<i> chain, with their "weighty, ponderous" vowels are all considered passive. Yet the activity scale shows less correspondance with Martin's scheme: although stimuli follow the same ranking as passivity, there is no chaining of the non-open vowels. On the brilliance scales, the "light" <a> is ranked as brightest/least-dark; this recalls Martin's explicit association between heaviness and darkness (Martin 1962: 184), and thus its corollary, that lightness connotes brightness. But the brilliance ranking of <i>, whose vowel should be either "gloomy," ("heavy") or neutral, is not as would be predicted, since it collapses with the light <a>. However, it is on the size scales that the data most clearly conflict with Martin's vowel scheme. The largest/least-small stimulus is <a>, according to Martin, a "petite, delicate, insignificant" vowel, while <i>, characterized as either "ample, weighty, bulky" or neutral, is considered least-large/smallest. Moreover, on both size scales, there is an <u/a> collapse; thus <a>, the only light stimulus vowel, is not distinguished from <u>, which is invariably heavy.

If Martin's light-heavy vowel distinction appears to be no more than partly borne out by the data, then the semantic neutrality of <i>, and possibly <ʌ>, is not supported too well either. Ranking evidence is inconclusive: <i> occupies a middle rank on 3/6 scales (activity, passivity and darkness) and <ʌ> on 2/6 scales (largeness and smallness). If neutrality is indicated by a tendency for the 'neutral' stimulus to collapse with its partners, then <i> appears the opposite of neutral, since is differentiated from its three pair-partners 83% of the time, on $15/18^{45}$ tests. As observed, <ʌ> collapses with <u> on 5/6 scales, but it is differentiated from the other two stimuli (overall

---

stimulus.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

differentiation for <ʌ>from the other stimuli is 12/18 tests, or 66.6%). This suggests that the reason for the <ʌ/u> collapse is unlikely to be <ʌ>'s semantic neutrality.

As far as this study is concerned, the most interesting thing about Martin's vowel scheme is its claim that jaw-openness/tongue position is the primary cue for a knot of phonosymbolic attributions in Korean. Such a viewpoint further supports our hypothesis that the unexpected endpoint brilliance and energy rankings of <a> in the Korean data are due to the language- or culture-specific foregrounding of <a>'s amplitude/openness. However, while Martin (1962: 177) argues that his findings offer counter-evidence to the idea of a universal phonosymbolism of size, his claims are not borne out by the Korean size data obtained in this experiment. On the contrary, the Korean size rankings agree so well with those of the AE subjects as rather to give evidence of universality. In addition, there are aspects of the brilliance and energy results which cannot be accounted for by Martin's symbolism.

Two further explanations can be offered here. Firstly, since the stimulus vowels produced by these subjects were most probably in an interlanguage mixture of AE and Korean,[46] subjects may not have felt able to categorize them—or categorize them unconflictedly, without reference to other features—according to their central first-language phonosymbolism of heaviness and lightness. And secondly, it is possible that Martin's categorization represents only one phonosymbolic dimension of Korean vowels, while others yet to be defined are responsible for the judgments of the Korean subjects in this experiment, especially those on the size dimension.

---

[45] This total represents 6 scales x 3 stimulus-pairs (in this case, <i/ʌ>, <i/a> and <i/u>).
[46] For discussion of this point, see Chapter 6, Section 6.2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Further experimentation should therefore include a) a heavy-light semantic scale, specifically for the testing of Korean phonosymbolic responses against Martin's claims and b) recordings of the pronunciation of the stimuli by a representative subject sample of each of the ESL groups, to clarify how far stimulus pronunciations diverge from the AE targets.

Salient versus nonsalient robustness of response. The scatterplot, Figure 8.5 below (which corresponds to Figure 7.6 in Chapter 7),[47] allows us to see whether the greater robustness found in the AE salient scale responses is also found in the Korean data.

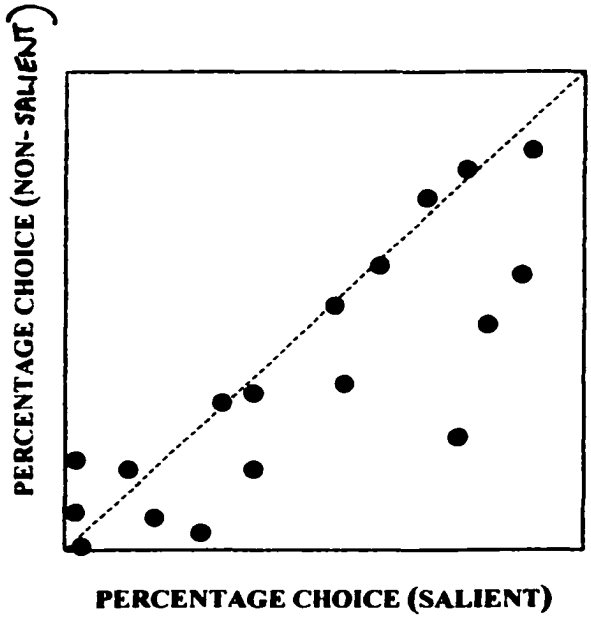


Figure 8.5: Preference strength in antonymic scales (Korean speakers). Percentage of elective choices ('neither' responses omitted) favoring preferred stimuli for each of 6 stimulus pairs evaluated on salient vs. non-salient sub-scales of 3 dimensions. Axis values run 50 to 100%.

[47] In this figure, as in the previous chapter's Figure 7.6, dimension by dimension, corresponding cells of salient and non-salient scales in the frequency matrices were examined, and the preferred stimulus in each comparison was plotted as a proportion of the elective choices which it received. These representatations of comparisons were made without regard to stimulus pair or dimension.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

At first glance, Figure 8.5 might appear to indicate a more even distribution of values than in the AE case. Although there are 10/18 instances of points below the diagonal (indicating numerically stronger salient scale preferences for a stimulus), there are 3/18 on the diagonal (preferences of equal strength on salient and non-salient scales) and 5/18 above (indicating stronger non-salient scale preferences). However, further inspection reveals that 7/10 salient points are located well below the diagonal, while 4/5 non-salient points are located closely above it. In other words, point positions relative to the diagonal indicate *smaller majorities* for the preferred stimuli on the 'winning' non-salient scales, and larger majorities for those on the 'winning' salient scales. Thus, it can be claimed that the tendency found in the AE data for a more robust response pattern on the salient scales is repeated here.

Triggering features. As in the AE data, the Korean rankings are the same for brilliance and energy, and thus, the questions raised in the AE case are relevant here also. Firstly, were brilliance and energy in some sense synonymous for the Korean speakers? Secondly, were the same triggering features operating in evaluations on both dimensions? In answer to the first question, a qualified "yes" would be the answer. Despite identical rankings (suggesting similar criteria of evaluation) in the upper parts of the scales, we see striking configurational differences between the dimensions. Also, saliency affects the dimensions differently: whereas ranges and configurations are similar on both brilliance scales, there are large range and configurational differences between the salient and non-salient energy scales. These interdimensional differences are most strikingly exemplified in the placement of <i>, which receives very different evaluations on the brilliance scales than it does on activity, and is differently evaluated yet again on passivity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

It therefore appears that although the same triggering features may have been operating, they did not carry the same weight on both dimensions. Rather than gradations of one dominant feature, it can be claimed that interactions between two features (or, more properly, between two feature-*gestalten*, linked sets of acoustic and articulatory features) are responsible for the rankings. On the brilliance dimension, anteriority can be seen as the main ranking criterion, but the <a/i> collapse and the endpoint ranking of <a> also indicates a *feature-equivalence*, with <a>'s amplitude/openness at least as as important as <i>'s high locus/anteriority. On the energy dimension, <a>'s independent endpoint position suggests that the amplitude/openness feature dominates high locus/anteriority, while their lack of amplitude/openness will explain the chaining of <ʌ>-<u>-<i> at the passive end of the non-salient scale. Finally, on the size dimension, although articulatory spaciousness can be considered as the main trigger, on the non-salient scale, we also see possible effects of other size-connoting features: perceived duration (in the differentiation of <ʌ> from <u>), shared lack of locus and/or lower F_0 (in the <ʌ/a> collapse).

Phonosymbolic profiles of the stimuli

<a> Non-consensus 5/18 (28%) This stimulus is found at an endpoint on 5/6 scales, and shares the endpoint rank step on the remaining scale, smallness. We see that <a>'s differentiability from other stimuli varies with dimension: on the energy scales, it is ranked independently, as most-active/least-passive, but as brightest/least-dark, it is not differentiated from <i> on the brilliance scales. It is also not differentiated from <u> on size. It can be described as a very bright, active (i.e., vital) and large vowel.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

<i> Non-consensus 3/18 (17%). This stimulus is independently ranked at an endpoint on 2/6 scales, as least-large/smallest. The capacity of <i> to be differentiated from other stimuli varies with dimension, or, in the case of energy, with saliency. As mentioned above, it shares the endpoint step with <a> on the brilliance scales; in contrast, on the size scales it is the only stimulus with an independent ranking. On the energy dimension, <i> is independently mid-ranked on activity but collapses with <u> on passivity, where its position in the <i>-<u>-<ʌ> chain might suggest a small bias in the direction of passiveness. In sum, it can be characterized as a fairly bright vowel, undetermined or neutral with regards to energy (perhaps with a tendency to passiveness),[48] and small.

<u> Non-consensus 8/18 (44%) This stimulus occupies a middle rank on all scales except that of smallness, where it shares an endpoint rank step with <a>. In these data, <u> is the least well-differentiated of the four stimuli; it invariably collapses with at least one rank neighbor and is also indeterminately placed on two scales from unrelated dimensions (largeness and passivity), that is, it twice participates in a double collapse. It is not differentiatied from <ʌ> on any scale except smallness. However, <u> can still be characterized as fairly dark and passive (i.e., non-vital), with a tendency towards largeness.

<ʌ> Non-consensus 6/18 (33%) This stimulus is ranked at an endpoint on 4/6 scales, those covering the vitality dimensions of brilliance and energy, while on the size scales it occupies a mid-range position. As mentioned above, it participates in an across-the-board collapse with <u>, but otherwise appears to have a well-differentiated phonosymbolic profile. It can be claimed

[48] Since for <i>, brilliance judgments ≠ energy judgments, no vitality generalizations are made about this stimulus.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that <ʌ> is a very dark and passive, non-vital vowel; it is either undetermined/neutral with regards to size, or (judging from its unique differentiation from <u> on the smallness scale) may have some tendency towards smallness.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 9

ANALYSIS AND INTERPRETATION OF THE JAPANESE DATA

9.1 The phonetic/phonemic background of the subjects

9.1.1 Articulatory details

The Japanese vowel system is simpler than those of English and Korean; it is generally held to be composed of five vowels.[1] It appears similar to the five-unit system of Spanish, but contrasts with it in two important ways: Japanese differs in the articulation of its high non-front vowel, [ɯ], and has one-mora and two-morae versions of its phones, thus doubling its repertoire. Japanese high vowels are subject to devoicing under certain conditions, while the other vowels have lesser, or no, propensity to do this.[2] Devoiceability must be considered as a potentially important feature of Japanese phonosymbolism, since it offers yet another way of vowel ranking; it can be added to those hierarchies produced by ordering vowels according to their intrinsic F_0's, durations or intensities.

---

[1] But see also Homma (1973: 350 and below), who asserts that there are six vowels in the Tokyo dialect.

[2] For details of devoicing, see below, Section 9.1.2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In a study contrasting English and Japanese vowels, Isshiki (1965: 196) describes the general articulation of Japanese vowels as non-extreme, "produced without moving the tongue too high nor too low in the mouth cavity...without opening the mouth very wide, or making the lips spread or round: they are usually lax sounds." Vance (1987) tends to agree with this description but provides more detail. His discussion of the phonemes of Japanese is preceded by an account of basis-of-articulation differences between Japanese and other languages, notably English (including British RP and his own mid-Western American). These differences are examined in terms of "articulatory settings...muscular adjustments involved in voice quality" (Vance 1987: 6) that are extrinsic, i.e., acquired by social imitation rather than developed idiosyncratically. Three parameters are involved: the movements and positions of the lips, tongue and lower jaw. The most important variation is in lip activity, Vance (1987: 7) noting that it can be significantly contrasted in Japanese, English and French, with French having the greatest degree of rounding, while rounding is "relatively weak in English" and "even weaker in Japanese." The openesss of jaw position is also contrasted; French has the widest opening, and Japanese is "distinctly more open" than English (Vance 1987: 7, disagreeing here with Isshiki). Tongue activity in English and Japanese also displays a basic contrast, English maintaining contact with the roof of the mouth through lateral tethering or "anchorage" while Japanese anchors the tongue "lightly and centrally on the floor of the mouth" (Vance 1987: 8).[3]

[3] This setting may facilitate the pronunciation of Japanese back segments, including [a]. However, Vance (1987: 8) remarks that "for some speakers this [vowel] is more central than back." Since AE and Korean [a] are also reported to vary in anteriority (see Chapter 7, Sections 7.1-2 and Chapter 8, Sections 8.1-2), this suggests that across languages [a] tends to be horizontally unstable.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The point vowels. There is some disagreement on the tongue placement of Japanese [i]. Isshiki (1965: 197) describes it as being articulated more centrally than its English counterpart and a little lower, and Akamatsu (1997: 30), whose focus is more narrowly phonetic, agrees with him, describing it as "less close and less front than CV 1" (cardinal [i]). On the other hand, Vance (1987) maintains that Japanese [i] resembles cardinal [i] in tongue position, which would yield a vowel with a higher, more forward tongue placement than the starting position for AE [iy].[4] Isshiki describes Japanese [i] as articulated with "loosely spread" lips, but both Vance and Akamatsu go further, claiming that, unlike French or German [i], or English [iy], it lacks the feature of lip-spreading: "The lips assume a neutral posture" (Akamatsu 1997: 30). Akamatsu also emphasizes the shortness and steady-state quality of Japanese [i], compared to its English equivalent. Finally, we note that Hamano's (1986: 97, 101,146) descriptions of the expressive function of [i] in Japanese suggest the phonosymbolic importance of a high, forward placement, rather than a lower, more central one: "A high-pitched sound...sharpness of sensation...whistle of a small bird...clinking...ringing...[giving a] meaning of thinness, sharpness, slenderness and tenseness."[5]

Vance describes Japanese [ɯ] as varying considerably from its French (or English) counterparts [u] and [uw]; as with [i], it is the partial or total absence of lip activity which distinguishes it.[6]

---

[4] I am informed here by Professor Edward Bendix (personal communication). See also Ladefoged (1982: 199), who describes cardinal [i] as having a "more extreme quality" than English [iy].

[5] Hamano's characterizations also suggest a high locus of concentration of acoustic energy for [i]; see below, Section 9.1.2.

[6] Vance specifies that what is lacking—except in very careful speech—is lip *compression* rather than rounding: "This compression is substantially weaker and sometimes totally absent in connected speech at normal

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Isshiki (1965: 203) describes Japanese [ɯ] as articulated a little lower than English [uw], with a slightly more forward tongue position; he also stresses its neutral lip position. We should note, however, that not every rendition of [ɯ] will be characterized by absence of lip gesture. Akamatsu (1997: 97) mentions [ɯ^{w}], a rounded or lip-compressed allophone of [ɯ],which is found following [Φ], the bilabial fricative allophone of /h/, and Professor William Stewart (personal communication) has observed that the rounding of Japanese [ɯ] is gender-associated, i.e., with masculine speech style. In addition, Hamano's (1986: 100, 102) descriptions of the expressive function of [ɯ] give evidence of the phonosymbolic importance of its rounding: "The event involves a small round opening...pushing out air or a small object through a small opening, such as...the protruded lips...[indicates a] small round shape...Smallness...accompanied by protrusion."

Akamatsu (1997: 31) agrees with Isshiki in locating Japanese [ɯ] below cardinals [u] and [ɯ] (CV's 8 and 16) and more centrally, characterizing it as a "close back vowel." However, he finds the tongue position of British and American [uw] "practially the same as for that in the articulation of Japanese [ɯ]... The vowel in food, mood, loot etc., *if well unrounded and shortened* would sound like the Japanese [ɯ]." (Akamatsu 1997: 31). Morimoto (1988), in a contrastive acoustic study of the vowels of peninsular Spanish and Japanese, agrees with Akamatsu and Isshiki, comparing [ɯ] with CV 16, but placing it more forward "in the area of central or centralized vowels" and enough below CV 16 to "neutralize its more closed quality" (Morimoto 1988: 189).

---

conversational tempo.... Although the compression overrides the articulatory setting in careful speech, the setting

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Homma (1973) made a detailed acoustic investigation of the vowels of her own speech and discovered that her [ɯ], already much more centralized than she had found described,[7] tended to further centralize when following coronal consonants. Thus, whereas acoustically the other four vowels were well separated from each other, her [ɯ] partly overlapped with [i] in its formant values (Homma 1973: 351-353). To accomodate this discovery, she added a sixth vowel, [ʉ̟], a fronted allophone of [ɯ], to her chart; significantly, none of her other vowels showed the kind of variants that would need allophonic 'extras' to represent them. Homma's study is thorough[8] and much-cited, yet since her subject sample was restricted to one, it raises questions about the representativeness of her findings. Akamatsu (1997: 32) partially supports Homma's observations in stating that the centralized [ʉ̟] will follow the apical fricative [s] and affricates [ts] and [dz],[9] but since for all the stimuli, the initial consonant was [h], this should not have conditioned the production of the centralized allophone. However, the use of prevocalic [h] in the stimulus syllables may produced further complications in pronouncing them;[10] this possibility will be kept in mind in the data interpretation sections.

---

begins to dominate in more rapid speech" (Vance 1987: 11).

[7] Homma's reference is the *Kokugogaku Jiten (Dictionary of Japanese Linguistics)* 1955, Tokyo: Tokyodo Press. She agrees with the Dictionary in the six vowels of the Tokyo dialect, but places both her /ɯ/ allophones more forward.

[8] Homma's findings are based on over 300 spectrograms (of 155 artificial words to investigate vowel formants, and of 154 disyllabic real words to test the effects of pitch, amplitude and duration on pitch-accent) .

[9] Homma found in her speech that the centralized allophone followed alveolar and palatal glides and nasals as well as [s], [ts] and [dz].

[10] According to Akamatsu (1997: 97), "[h] does not occur before [i] or [ɯ] in Japanese," assimilated variants being found instead: the voiceless palatal fricative [ç] before [i], and, as mentioned above, the voiceless bilabial fricative [Φ] before [ɯ]. I am also informed here by Dr. Edward Bendix, and by Professor William Stewart (personal communications). For allophones of Japanese [h], see also Han (1961: 37).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In Vance's description, Japanese [a] occupies a low, slightly backed but basically central position, more or less identical to its location on Ladefoged's (1982: 204) chart of the Japanese vowels. Although Vance does not take up this point, if the articulatory setting for Japanese does indeed differ from the English in having more open jaw position, then Japanese [a] should be lower than its English equivalent. Vance does, however, raise questions about the anteriority of Japanese [a]: "the range of individual variation... is rather wide," with one (female) authority[11] giving it "a quite front character" (Vance 1987: 10). This may be partly gender-determined: Professor William Stewart (personal communication) has observed that fronting of vowels in Japanese is associated with female speech, which agrees with Homma's account of her own [a] and [ɯ]. Further evidence for this view is provided by Han (1961: 7) in a study based on 300 spectrograms of the speech samples of one male and one female radio announcer; she found that the female's articulation was "much more toward the front than the male speaker."

Isshiki (1965: 195, 202) identifies English [a] as a "low-central vowel," and Japanese [a] with the first element in the diphthong of *try*, its main area of lingual tension being "center... [to]...center of back." Akamatsu (1997: 34) places Japanese [a] more or less exactly in the center of the vowel quadrilateral, midway between the cardinal vowels 4 and 5; he remarks that its "quality may simply be said to be open. The mouth is fairly open, as the jaw is reasonably lowered... (It) will not present difficulty to native speakers of Spanish or Italian, as it resembles their own."

---

[11] This is Homma, who describes her own [a] as a "near front" vowel (Homma 1973: 253).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

However, Morimoto (1988: 187) asserts that the tongue position is farther back for Japanese than for Spanish [a].

Akamatsu (1997: 35) gives diagrams of his own and two other versions of the tongue positions for the vowels relative to the cardinal points. These illustrate the horizontal instability of [a] and the tendency of [ɯ] to centralize. He observes that "the five Japanese vowels are in general somewhat inside the periphery of the cardinal vowel area, unlike e.g. English vowels, at least half of which are on the periphery of the cardinal vowel area, or French ones most of whose vowels are on the periphery". He re-emphasizes the stable, monophthongal nature of all the Japanese vowels: in this respect, Japanese resembles Spanish as it does in its five-vowel system.

To sum up, we find several important contrasts in comparing the AE target versions of the stimulus (point) vowels with their closest Japanese equivalents. Firstly, Japanese [a] resembles AE [a], in its openness and tongue position (central or backed-central), although its basis of articulation is different. These subjects' <a> therefore should have resembled those of the AE and Korean groups, with possibly a greater degree of openness. Secondly, Japanese [i] differs from AE [iy] in having little or no lip spreading; tongue positions of the two vowels are similar, but Japanese [i] is probably slightly lower and less peripheral. We might therefore expect that this group's <i> would be less extreme in articulation than those of the AE and Korean subjects, and that this might be reflected in the evaluations. Thirdly, Japanese [ɯ] contrasts with AE [uw] in its lack of rounding; it is likely to be articulated with a tongue position which is farther forward and a little lower than its AE equivalent. (We will reserve accounts of how the Japanese subjects are expected to have pronounced <u> for the following discussion on their likely

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

renditions of <ʌ>.) Lastly, Japanese male and female articulations of the two nonanterior vowels may differ horizontally, being articulated with more forward tongue positions by females; the lip rounding of [ɯ] may also vary between the sexes, with males rounding more than females. Gender-linked differences of this type are not noted in AE vowels.

Japanese vowels corresponding to AE [ʌ]. In the previous chapter (see Sections 8.1.1 and 8.1.2), it was argued that Korean subjects could have perceived any of several unrounded non-front L1 vowels as the one which most closely corresponded to AE [ʌ], and thus as their underlying references for pronouncing the stimulus vowel <ʌ>. Accounts of the Korean vowel inventory indicate sufficient items to provide a phonemic referent for <ʌ> separate from those for <a> and <u>; however, the cluster of <ʌ/u> collapses in the Korean data suggests confusion, possibly pronunciation-generated, between these stimuli (for discussion, see Sections 8.1.2 and 8.2.1). If <ʌ> gave problems even for subjects who were considered to have an L1 equivalent to AE [ʌ], then difficulties with this stimulus can be predicted for the Japanese subjects, who had a more limited vowel inventory, and greater likelihood of conflicts in choosing an L1 phone as a reference for <ʌ>.

Out of the five vowels in the Japanese repertoire, [i] and [e] should have been ruled out as anterior; [o] should also have been excluded as rounded, although the orthography of four items on the word-list might have influenced pronunciation.[12] Only [a] and [ɯ], are unrounded and non-front, but the choice of either as an AE [ʌ]-equivalent and referent for <ʌ> would have posed

[12] These were spelled with *o* or *ou* (mother, some, other, couple). It is possible, though not likely, that they were pronounced with [o], since they are commonly occurring words, i.e., [ʌ]-pronunciations would have been familiar.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

a problem, since L1 pronunciation reference for three AE stimulus vowels, <a> <u> and <ʌ>, would have had to be provided by the two Japanese non-front unrounded vowels. We can explore both options for <ʌ> here, but we cannot definitively answer the question of which was the more probable choice until the Japanese data are analyzed (see Section 9.2). We can, however, predict the most likely patterns of stimulus differentiation and collapse resulting from the selection of either [a] or [ɯ], and subsequently match predictions to the actual findings.

Isshiki (1965: 202) asserts Japanese [a] is the vowel perceived as most closely corresponding to AE [ʌ]: "As there is no stressed or unstressed [ə] in Japanese, Japanese speakers tend to substitute [a] for this... Japanese speakers tend to confuse [a] [and] [ə]." In a personal communication, Ms.Takako Katsuno of the CUNY Graduate Center Linguistics Department[13] supported this view; she assured me that for Japanese speakers, [ɯ] was not a likely referent for English [ʌ], and that [a] was the obvious choice, pointing out that in the English loanwords *uptown* and *subway*, the vowel in the first syllable is pronounced as [a] by Japanese speakers. Two-thirds of the <ʌ> words on the word-lists (including subway), were spelt with *u*;[14] it is possible that subjects (besides being familiar with their pronunciations and thus remembering that these should be pronounced with the closest equivalent to [ʌ]) were also influenced by their orthography to render them with Japanese [a]. However, Ms. Katsuno also remarked that where [ʌ] is represented by *o* as in *London*, Japanese pronunciation will be with [o]. As mentioned, the other third of the word-list spellings were with *o*; there is a small possibility this may have

[13] Ms. Katsuno was the phonetically sophisticated Japanese informant who back-translated the Japanese subjects' responses to the fourth page of the questionnaire (their translations of the semantic scale adjectival terms).

[14] These were fuss, subway, but, teacup, accustomed, rush, dusty and under.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

suggested the [o]-association, at least for these words, or otherwise rendered the pronunciation of the stimulus <ʌ> less consistent.

Solving the problem of <ʌ>. If Japanese subjects chose [a] as their vowel most closely corresponding to AE [ʌ], then its parent phoneme, /a/, would have had to do a double service, covering both <a> and <ʌ>. One solution would have been to pronounce both <a> and <ʌ> as the same allophone [a], represented as follows: <a>←[a]←/a/→[a]→<ʌ>.Given the experimental task of differentiating between the <ʌ/a> pair, this appears highly unsatisfactory, but if we find in the Japanese results a cluster (five or more) of <ʌ/a> collapses, then we might have evidence that this strategy had been employed. A second and better solution would have been to have used two different /a/ allophones, ensuring more separable pronunciations for <ʌ> and <a>. But /a/ (consistently characterized as a low tongue position/open vowel with no marked lip gesture) would seem a less than perfect candidate for dividing in two in this way, because its allophones will vary only along the parameter of horizontal variability.[15] This solution to the problem of <ʌ> might be tentatively represented as follows: <a>←[a˂]←/a/→[a˃]→<ʌ> (where [a˂] and [a˃] represent backed and fronted /a/ allophones). However, following this second strategy, the two stimulus vowels could still have seemed less than perfectly differentiable; once again, we would predict the likelihood of several <ʌ/a> collapses in the data.

Although a third solution, selecting [ɯ] as the L1 reference for <ʌ>, might initially appear less likely, there was a substantial advantage to be gained by it. The phoneme /ɯ/ has two allophones,

[15] For articulatory accounts of [a], see above. We note that there are no descriptions of specifically-distributed /a/ allophones, as there are in the case of /ɯ/.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

a fronted [ʉ] and a rounded, probably backed[16] [ɯ^{w}] (see Homma (1973), Akamatsu (1997), and the discussion above); variation would be along two parameters, and thus subjects could better maintain an articulatory and psychological separation between renditions of <ʌ> and <u>.We recall how the target stimulus vowels were presented for learning in the early part of the experimental procedure. The method for establishing their satisfactory pronunciation consisted of each subject's reading down the word lists and, after some individual practice, pronouncing list items aloud in unison with the others in the room. There was pressure to conform one's pronunciation to the approved norm, since satisfactory and unsatisfactory (misread or obviously mispronounced) renditions of the words were signaled by the experimenter with either a nod or a shake of the head. All that was needed were one or two subjects whose renditions were audible and obviously approved by the experimenter for the whole group to imitate them—thus standardizing pronunciations.

In practicing with the <u> list, the experimenter would have encouraged subjects to produce a vowel as close as possible to a well-rounded [uw] by nodding with approval to those who were clearly doing so. It is also likely that many of the Japanese speakers were experienced enough in English pronunciation[17] to know that they needed to add lip rounding (or lip compression) to produce an English-sounding [uw]. Corresponding to Vance's (1987: 11) "careful speech," in the slow, focused pronunciation of isolated English words (as on the word lists) or syllables (as on the questionnaire itself), rounding/lip compression on [ɯ^{w}] would be a relatively easy thing to

---

[16] Some tongue retraction can generally be expected to accompany lip rounding. I am advised here (personal communication) by Professors Lawrence Raphael and Edward Bendix, author of the Bendix Method of Japanese Transcription, in use at Columbia University.

[17] Ninety percent of the Japanese subjects had been studying English for over two years (see Appendix 2).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

maintain consciously, much easier than in connected speech. In practicing with the <ʌ> list, subjects with an obvious tendency to substitute Japanese [a] for English [ʌ] would have been discouraged with a head-shake; the experimenter did not so discourage substitution of Japanese [ɯ] (or the more centralized [ʉ̟]) for English [ʌ], since to an (R.P) English-speaking ear, they seemed much closer to this vowel than is Japanese [a]. In addition, the lists for <u> and <ʌ> were found one above the other on the same page.[18] Thus, even before the stimuli were paired for evaluations, besides external pressure from the experimenter, subjects may have experienced an internal need to keep the two vowels articulatorily as distinct from each other as they could. This would argue for the most polarized pronunciation of <ʌ> possible, using the allophone farthest away from [ɯ^{w}], that is, [ʉ̟] rather than [ɯ].

There are two last points in favor of the <u>-as-[ɯ^{w}] hypothesis. Firstly, if <u> were pronounced with a rounded vowel, the rounding may have been reinforced for Japanese subjects by the stimulus spelling *hoo*, since [o] in Japanese is the only consistently rounded vowel. Secondly, to have added rounding to one of the stimuli also meant that although [o] was not part of the experiment, its features of rounding and backing were still not excluded. In this way, crucial Japanese phonetic/phonological contrasts normally thrown into relief by the presence of [o] (with the anteriority and unroundedness of [i], the openness, unroundedness and non-backness of [a]) could continue to function in its absence. Thus, these contrasts would be available for phonosymbolic use in responding to the questionnaire.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This third strategy for pronouncing <ʌ>can be represented as <ʌ>←[ɯ̈]←/ɯ/→[ɯ^{w}]→<u>. If it were the one which subjects used, we would predict that, unlike the Korean case, the <ʌ/u> pair should stay well differentiated over all six scales. However, the <ʌ/i> pair might prove more liable to collapses, since, as Homma (1973) observes, [ɯ̈] and [i] tend to overlap phonetically, and the unusual experimental environment may have brought this out.[19] In other words, the cost of the strategy aimed at keeping <u> and <ʌ> separate might turn out to have been borne by <i>.

9.1.2 Vowel devoicing

Devoicing is triggered in two environments: interconsonantally, where both consonants are voiceless, and word-finally preceded by a voiceless consonant (Han, 1961).[20] The predisposition of Japanese high vowels to devoice, a phonological/phonotactic feature, appears to be conditioned by several factors. One is tongue height/jaw openness, which is articulatory, and another, intrinsic duration, which is primarily acoustic (though it may have an articulatory component).[21] A third factor is pitch accent, which is acoustic. Since our discussion thus covers both articulatory and acoustic phenomena, vowel devoicing is treated here in a separate section.

---

[18] Lists for <a> and <i> were presented together on the first page, and those for <u> and <ʌ> on the second. (See Appendix 1for the handout with the four word lists.)

[19] "Unusual" here would mean attempts at an AE stimulus-initial [h]. I am advised that in their normal Japanese environments, [i] and [ɯ̈] although overlapping in their phonetic ranges, would be in complementary distribution, and thus kept psychologically separate. That is, the [ɯ̈]allophone of /ɯ/ occurs after the allophones [ts], [dz], [s], [(d)z] of /t/, /d/, /s/, /z/ before /ɯ/, whereas the [i] allophone of /i/ occurs after the allophones [ç], [j hook], [ʂ], [j hook] of /t/, /d/, /s/, /z/ before /i/. Thus, in Japanese [i] and [ɯ̈] can overlap phonetically, but the particular allophone of the preceding coronal cues whether the vowel is /i/ or /ɯ/ (Professor Edward Bendix, personal communication).

[20] The second of these devoicing rules is also true of some varieties of Italian, Portuguese and Spanish (e.g., Dominican).

[21] Although intrinsic duration is held to belong to the acoustic study of vowels, clearly, it could also overlap the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Akamatsu (1997: 36) asserts that [a], like [i] and [ɯ], is a short vowel, and that all five one-mora vowels are comparable in length, and contrast only with their two-morae versions. However, in a spectrographic analysis of the phonemes of three speakers of standard colloquial Japanese, Han (1961: 17) found inherent vowel durations to be different, the relative ratio among the vowels being [ɯ] 1.00, [i] 1.17, [o] 1.26, [e] 1.37 and [a] 1.44.[22] Moreover, she found the devoicing of vowels to be sensitive to durational differences, with "the relatively short vowels [ɯ] and [i]... more readily unvoiced than [o, e, a]" (Han 1961: 23). Homma (1973) used self-recorded disyllabic words and measured durations of both V_1 and V_2. She found her [a] to come in the middle of both durational series. with the mid vowels being longer in both cases and the high vowels shorter. Homma also quotes a study by Okada who, like Han, found [a] to have the longest duration, [ɯ] the shortest and [i] the next shortest, but does not mention how many informants Okada's study was based on.[23]

Isshiki provides more evidence of a relation between devoicing and vowel shortening, and for the high vowels, particularly [ɯ], as prime candidates for devoicing: "The Japanese vowels, especially [i] and [ɯ] are often devocalized...*especially when they are spoken rapidly*," i.e.,

---

articulatory area, since speakers may well include duration among the kinesthetic-proprioceptive sensations associated with the production of a vowel.

[22] Interestingly, according to Han, front vowels are longer than the back vowels at the same corresponding height, suggesting a potential feature conflict on the size scales, if both perceived length and backness correlate with magnitude. Following Sapir (1929) one would expect that [i] and [e], being front vowels, should be judged smaller than [u] and [o].

[23] Homma agrees with Okada in finding that her own vowels were longer when preceded and followed by voiced consonants; if there is a correlation between shortness of vowel duration and liability to devoice, then it is likely that voiceless consonants will further shorten a vowel with lesser intrinsic duration and thus predispose it for devoicing.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where tendencies to vowel contraction would be maximized (Isshiki 1965: 196, emphasis added). All five of Isshiki's examples of devoicing are of [ɯ] alone, in medial or final position, suggesting that this is the more likely of the two high vowels to devoice. Epenthesizability, an additional phonotactic/phonological feature of [ɯ],[24] may also contribute to its sense of minimal duration, and thus especial predisposition to devoice.

Further evidence that devoicing is linked to shortness of vowel duration has been provided by Dr. Winifred Strange (personal communication) who assured me that it is limited to one-mora vowels only. On the other hand, Professor William Stewart argues that the phenonenon is only tangentially related to mora-length.[25] Professor Stewart proposes pitch accent as a more likely cause of vowel devoicing, observing that only vowels with low pitch accent devoice, as in, e.g., main Tokyo dialect's *hashi*↑ 'bridge' vs.*hashi*↓, 'joss-stick', where devoicing occurs on the second word but not the first. He explains the predisposition of the high vowels alone to devoice as due to their closer articulation; when devoiced, [i] and [ɯ] will still leave a glidelike trace,[26]

---

[24] In loanwords, [ɯ] has an epenthetic function, serving to separate consonant clusters; it may also be inserted between [Φ] and a front vowel (Akamatsu 1997: 87) but this seems to occur in attempts to pronounce [fi] in loanwords, where the epenthesis of [ɯ] facilitates the production of the closer [Φ] allophone of /h/, for [f] before [i] rather than the [ç] allophone of /h/, which in earlier generations' attempts to pronounce [fi] came out as [çi] as the closest equivalent. Between consonants in loanwords, [ɯ] seems to operate like English [ə] in emphasis pronunciations of the initial clusters of *Christ* [kərayst] or *please* [pəliyz], as the shortest possible vocalic element which will ensure separation of the consonants. Another vowel, [o] also is used epenthetically (Professor William Stewart, personal communication), but [o] was not represented in the stimuli. Epenthetic [o] rather than [ɯ] is inserted after /t/ and /d/ in loanwords; [o] preserves the stop pronunciation, since [ɯ] would condition the affricated allophones [ts] and [dz] (Professor Edward Bendix, personal communication).

[25] Professor Stewart claims (personal communication) that the constraint on two-morae vowels' devoicing is due their being interpreted as V_1V_2 sequences, in which V_2 is reinterpreted as a voiced consonant, thus disqualifying the apparently acceptable sequences CV_1V_2# and CV_1V_2C (where the consonants are [-vce]) from triggering devoicing.

[26] It may also leave compensatory moraic lengthening of preceding fricativity (Professor Edward Bendix, personal communication). Thus /tɯ/ can go to [tsɯ̥] where [ɯ̥] is devoiced, and then to [tss], where the second [s] is syllabic.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

more easily recoverable from the rapid speech stream than the more open [e], [o] or [a] which will not leave such consonantal traces.

Since Professor Stewart's account of devoiceability contrasted so notably with those of the other authorities consulted, I asked a Japanese linguist, Dr. Takehiko Maruyama, of the ATR Spoken Language Translation Research Laboratories, about his impression of the phenomenon as a sophisticated native speaker. In a personal communication, he responded by confirming that [i] and [ɯ] were indeed most likely of the vowels to devoice, but that [a] and [o] were also capable of devoicing in specific environments ([ka], [ko], [ha], [ho], in initial position).[27] Since these are non-front vowels preceded by back consonants, this suggests that the posteriority of segments and their position in the word are yet other factors which may influence devoicing. Dr. Maruyama also gave as his impression that high pitch accent tended to preserve a vowel in the voiceless C_C environment from devoicing, whereas low pitch accent made it more likely to devoice: "High pitched syllables are pronounced strongly, so devoicing does not occur easily; low pitched syllables are pronounced weakly and devoice easily." On the other hand, he cautioned that devoicing "is not a necessary phenomenon, but...depends on time, circumstance and the speaker."

To conclude, from the above accounts, the liability to devoice, which could be described as a phonetic 'tendency-to-invisibility' feature, seems to be shared by [i] and [ɯ], possibly with [ɯ] the more predisposed to 'disappear' in this way. Of the separate influences playing into

---

[27] Could this have affected the pronunciation of <a> (phonetically [ha#])? Because the stimulus syllables were isolated, and not inserted into words, this would not appear likely.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

devoicing, it appears that high pitch on a vowel preserves 'visibility,' in contrast with low pitch; longer intrinsic duration and two-morae length preserve 'visibility,' in contrast with shorter intrinsic duration and one-mora length. I would also hypothesize, given Dr. Maruyama's remarks, that anteriority may preserve 'visiblity,' in contrast with posteriority. If [ɯ] is indeed more predisposed to devoice than [i], this could be due to its shorter duration and the epenthesizability associated with it, and perhaps, its more posterior position. Their shared predisposition to devoicing oppose [ɯ] and [i] to [a], which is less likely to devoice; this is an opposition which may have been used by subjects in making phonosymbolic attributions. Their devoicebility might have provided an extra feature for associating (i.e., collapsing) the high vowel stimuli in a triggering semantic environment; for example, we might see a collapse of <i/u> or <i/ʌ>, and an opposite ranking for <a>. On the other hand, differing predispositions to devoice may have helped subjects differentiate between <i> and <u> (or <i> and <ʌ>) in contexts where these pairs would otherwise have tended to collapse. [28]

---

[28] Our argument here is that the underlying phonemic contrast for both <u/i> and <ʌ/i> may have been /ɯ~i/. This is based on the hypothesis (see above, Section 9.1.1) that pronunciations of <ʌ>and <u> were both derived from /ɯ/-allophones, respectively [ɯ̟] and [ɯ̟ʷ].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

9.1.3 Acoustic details

A classic vowel study like Peterson and Barney's (1952), using a substantial number of subjects, indicating values for F_0 through F_3 and specifying formant intensities, apparently does not yet exist for Japanese. A number of sources of data have, however, been located; these will be used to build up a tentative acoustic account of Japanese vowels, covering formant values, intrinsic pitch, duration and intensity. It should be born in mind that all but one[29] of these studies provide data from a very limited number of informants, and should therefore be considered as no more than probable indications of the acoustic dimensions of Japanese vowels.

Kasuya, Suzuki and Kido's study[30] (1968: 359) compares the values for F_1 through F_3[31] of the vowels of between four and seven speakers in each of four categories: children under 12, youths (sex unspecified, but presumably male), adult men and adult women. Homma (1973: 352) provides a spectrographic study of her own vowels, giving the average values for F_1 and F_2; this includes two allophones of /ɯ/, [ɯ] and the fronted variant, [ɯ̟]. Average values for F_0 "occurring at the high pitch accent" (Homma 1973: 356) are also given, as are the amplitudes and durations of the vowels. In *A Grand Dictionary of Phonetics*, Onishi (1981: 201) offers the reader a spectrograph of all five Japanese vowels (i.e., of a single person's speech, sex

[29] In one instance (Onishi 1981: 672) the number of informants is unspecified.

[30] This study was in Japanese. I am indebted to Ms.Yuki Hirose of the CUNY Graduate Center Linguistics Department for her help in translating the relevant sections.

[31] Values for F_0, are not included, although the authors conclude that besides F_1 and F_2, either F_0 or F_3 "is an *indispensable parameter for the discrimination of vowels*, regardless of the age and sex of the speaker" (Kasuya, Suzuki and Kido 1968: 356, my emphasis). This conclusion, like Halberstam's (1998), suggests that both F_0 and F_3 may turn out to be important phonosymbolically.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

unspecified but from the values obtained, almost certainly male).[32] In the same work, Onishi (1981: 672) provides a table of F_1 and F_2 values, comparing Japanese vowels with those of French and English (number of informants and sex unspecified, but again, judging by the values, almost certainly male). Morimoto (1988: 185) compares acoustic analyses of Japanese and peninsular Spanish, using speech samples from five male and five female informants for each language; his study gives tables of F_1 and F_2 values for all four groups. Finally, Dr. Winifred Strange has very kindly provided data from a vowel study shortly to be published. Using four male speakers, she measured the formants (F_1 through F_3) and durations of one-mora and two-morae length vowels in the first nuclei of [hVba] bisyllables.[33]

Comparison of Japanese formant values drawn from six sources. Tables 9.1-3 below compare F_1 and F_2 values across the studies which have been described. Female formant values have been included for Japanese, in contrast with the equivalent AE and Korean sections; this is to permit a tentative exploration of Professor Stewart's remarks that gender-based differences may be found in the articulation of Japanese non-front vowels (i.e., a tendency towards anteriorization for females and towards backing/rounding for males). Focus on female values also allows for the inclusion of Homma's study, which uniquely presents two allophones of the high non-front vowel, [ɯ] and [ɯ̟]. Tables 9. 1-3 include the formants of the mid vowels [e] and [o] as an aid in comparison of height and anteriority.

---

[32] Given its context in the *Grand Dictionary*, before dismissing this portion of data as too small to be useful, we should bear in mind that it was probably selected as having representative value, as presumably, was Homma's.

[33] Formants were measured from samples taken at the 25%, 50% and 75% stages of the durations of the vowels investigated. Values which will be cited here are those taken at the 50% sampling, i.e., those farthest away from, and so least likely to be affected by, consonantal transitions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9.1: Japanese female F_1 and F_2 values (in Hz), drawn from three studies

	F_1 values			F_2 values		
	Kasuya et al.	Homma	Morimoto	Kasuya et al.	Homma	Morimoto
[a]	888	1046	990	1363	2075	1600
[e]	483	655	580	2317	2209	2510
[i]	325	354	320	2725	2886	2910
[o]	483	659	610	925	977	1010
[ɯ]	375	375	450	1675	1730	1590
[ɯ̟]	--	363	--	--	2187	--

Adapted from Kasuya et al. (1968: 359), Homma (1973: 352) and Morimoto (1988: 185)

Female values compared.[34] All three F_1 values for [i] indicate these vowels are of equivalent height/closeness. Similar F_1 values for [ɯ] and [ɯ̟], indicating the same degree of height/closeness as for [i], can be observed for two out of the three studies; the exception is Morimoto's [ɯ], which appears to be articulated somewhat lower than his [i] (close in values to Kasuya's mid vowels, although his own and Homma's [e] and [o] are lower still). F_1 values for [a] indicate that Homma's is the lowest/most open in articulation, and Morimoto's is similar, while Kasuya et al.'s suggest a less open vowel.

F_2 values for [i] indicate a definite difference in anteriority between Morimoto's (suggesting the most front articulation) and Kasuya's (the least), with Homma's [i] appearing close in anteriority to Morimoto's. Comparison of the F_2 values for [ɯ] and [ɯ̟] indicates the exceptional fronting of Homma's [ɯ̟], which is much farther forward than any of the [ɯ]'s; it appears to be of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

equivalent anteriority to her [e]. The non-back/unrounded quality of [ɯ] is clearly seen when its F_2 values are compared those of [o], which are much lower in every case, indicating [o]'s greater backing/rounding. Overall, [ɯ] can be described as a central, or backed-central, vowel; Homma's [ɯ] is markedly farther forward than Morimoto's, while Kasuya et al.'s [ɯ] is mid-way in anteriority between those of the other two.

If F_2 values for [ɯ] and [ɯ̟] are compared with those of [a], they indicate that Homma's [ɯ̟] and [a] are of equivalent anteriority: both of them appear to be front, rather than central, vowels. Morimoto's [ɯ] and [a] are further back but also of equivalent anteriority to each other. Kasuya et al.'s [ɯ] is forward of their [a], but Homma's [ɯ] is either considerably farther back of her [a], or shows an F_2 difference due to lip rounding. Compared among themselves, the wide range in F_2 values for [a] indicate considerable variation, with Homma's the farthest forward and Kasuya et al.'s the most back; each of the three study's [a]'s appears to be in a distinct horizontal location. In all three cases, the difference between F_2 values for [a] and [o] indicate either that [a], like [ɯ], is either well forward of [o] or that there is a contrast in lip-protrusion/rounding between these vowels. All of the acoustic details noted here agree with the articulatory descriptions quoted above, in Section 8.1.1: i.e., that there may be some variation in the height/closeness of the high vowels, but [ɯ] is central and [a] is horizontally variable. The only disagreement is seen in those F_1 differences indicating apparent variations in the openness of [a]. Since [a] is the most open of the Japanese vowels, and its F_1 is therefore the lowest whatever its value, group-to-group

[34] In this and all subsequent comparisons, the rule-of-thumb of Dr. Lawrence Raphael's (cited in Chapter 8, section 8.1.2) will be adhered to. Only those discrepancies in formant values of more than 75 Hz (for F_1) and more than 125 Hz (for F_2) will be held to represent real differences in the articulation/timbre of the vowels discussed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

differences here may reflect some variation (socially or geographically conditioned) in the bases of articulation of the speakers examined.

Table 9.2: Japanese male F_1 values (in Hz), drawn from six studies.

	Male F_1 values					
	Kasuya et al.	Onishi spectro-gram	Onishi group	Morimoto	Strange et al. 1-mora	Strange et al. 2-morae
[a]	775	750	600	760	618	716
[e]	475	542	450	480	437	443
[i]	263	375	250	310	317	295
[o]	550	458	450	480	430	423
[ɯ]	363	417	300	380	349	324

Adapted from Kasuya et al. (1968: 359), Onishi (1973: 201, 672), Morimoto (1988: 185) and Strange (1998).

Male values compared. F_1 values for [i] fall within the same general range, indicating a close/high vowel, with the exception of that taken from Onishi's spectrogram, which suggests a slightly lower/more open articulation. If F_1 values for [ɯ] are compared with those for [i], we find that in every case, [i]'s F_1 is higher than [ɯ]'s, suggesting a consistent tendency towards a slightly lower articulation for [ɯ] than [i]. Clearest evidence for this can be seen in the data of Kasuya et al., and possibly those of Morimoto, but in the other cases the discrepancy between the two values is well under 75 Hz., and therefore may not indicate anything of note. The Onishi spectrogram's F_1 values for [ɯ], as for its [i], indicate a lower articulation (comparable with some of the values for the mid vowels in other studies). For [a], there are two groupings of F_1 values: one (Strange, one-mora vowels; Onishi group) suggests a less open articulation. The remaining studies' F_1 values indicate a fully open/low tongue position for [a].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9.3: Japanese male F_2 values (in Hz), drawn from six studies.

	Male F_2 values					
	Kasuya et al.	Onishi spectro-gram	Onishi group	Morimoto	Strange et al. 1-mora	Strange et al. 2-morae
[a]	1163	1375	950	1320	1182	1175
[e]	1738	1500	1900	1960	1849	1982
[i]	2263	2208	2220	2360	2077	2243
[o]	838	917	750	870	805	730
[ɯ]	1300	1250	900	1320	1171	1160

Adapted from Kasuya et al. (1968: 359), Onishi (1973: 201, 672), Morimoto (1988: 185) and Strange (1998)

A range of fairly similar F_2 values for [i] all indicate a front articulation, with only one (Strange, one-mora vowels) possibly more central. If F_2 values for [ɯ] are compared, we find greater variability, indicating three distinct horizontal locations for this vowel. A backed position is suggested by the Onishi group table value, and a more anterior position by those values measured from Onishi's spectrogram, found in Kasuya et al.'s and in Morimoto's data. A separate, backed-central, location between these two is indicated by both of Strange's values for [ɯ].

As with the female data, when F_2 values for [ɯ] and [a] are compared, we find that their pronounced similarities indicate equivalent horizontal positions for both vowels in the majority of cases: here, four out of six. These are Strange's two data sets (backed central), those of Morimoto (more anterior), and the group figures from Onishi's table (more backed). F_2 values for [ɯ] and [a] from the other two sources fall just outside the 125 Hz limit of similarity; they indicate that [a] is back of [ɯ] for Kasuya et al., but anterior to [ɯ] in the values taken from

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Onishi's spectrogram. Also as with the female F_2 values, [o] appears to be the vowel with the backmost/most lip rounded articulation, with [ɯ] and [a] anterior to it in every case. Finally, comparison of F_2 values for [a] suggest three different horizontal locations for this vowel, ranging from a more back articulation (Onishi's group figures) to a more central one (Morimoto's data and Onishi's spectrogram). Again, as with the female formant values, we note that the acoustic picture agrees with the articulatory one given in the previous section, except for F_1 differences for [a], indicating variations in its openness. The same explanation (that group-to-group differences in openness may reflect social or geographic variations in bases of articulation) is offered here.

Differences in anteriority of non-front vowels: evidence for a male-female difference? Our conclusions must be extremely tentative in this area, since the numbers of informants per study are small, and there are fewer data available for females than males. We have already noted that regardless of gender, [a] and [ɯ] are often articulated in the same place horizontally for the same speaker group (in six out of nine data sets), but positions for those vowels vary from group to group. If F_2 values are compared within each of the female studies, we find only in one out of the three, Homma's, that there is clear evidence for an anteriorizing tendency: the values for [e], [ɯ̟] and [a], indicate a similarly anterior placement for all three vowels. The other female F_2 values for non-front vowels give no especial evidence for a tendency toward front articulation. Homma's (self-described) central [ɯ] is close in F_2 value to Kasuya et al.'s, [ɯ] and slightly anterior to Morimoto's central [ɯ] and [a], while Kasuya's [a] can be described as backed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Comparing male F_2 non-front vowel values, we note that in four out of the six studies there is some evidence for a tendency toward a more back articulation. F_2 values for [ɯ] and [a] in Onishi's group table indicate a backed position, while a trend towards backing may possibly be seen in both sets of Strange's F_2 values for [ɯ] and [a], and in a similar value for Kasuya's [a]. On the other hand, Morimoto's F_2 values for [ɯ] and [a], and the F_2's seen for [ɯ] in Kasuya et al.'s data and measured from Onishi's spectrogram suggest central vowels. We conclude that gender-based fronting and backing tendencies may indeed exist for Japanese speakers, but that F_2 comparisons from the studies quoted above yield insuffficient evidence firmly to support Professor Stewart's observations. However, (assuming that similar tendencies are not found in American English non-front vowels) it is possible that in the comparison between male and female Japanese and AE formant values taken from Peterson and Barney's (1952) study, this issue can be further clarified.

Japanese formant values compared with those for American English. Tables 9.4-5 and 9.6-7 compare F_1 and F_2 female and male values of the point and mid vowels from Peterson and Barney's (1952) study with the Japanese values quoted above.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9.4: Comparison of female F_1 values (in Hz), drawn from three Japanese and one American English study[35]

Ja.	Kasuya et al.	Homma	Morimoto	AE	Peterson & Barney
[a]	888	1046	990	[a]	850
[e]	483	655	580	[ɛ]	610
[i]	325	354	320	[iy]	310
[o]	483	659	610	[ɔ]	590
[ɯ]	375	375	450	[uw]	370
[ɯ̟]	--	363	--	[ʌ]	760

Adapted from Kasuya et al. (1968: 359), Homma (1973: 352) Morimoto (1988: 185) and Peterson and Barney (1952: 183)

Table 9.5: Comparison of female F_2 values (in Hz), drawn from three Japanese and one American English study.

Ja	Kasuya et al.	Homma	Morimoto	AE	Peterson & Barney
[a]	1363	2075	1600	[a]	1220
[e]	2317	2209	2510	[ɛ]	2330
[i]	2725	2886	2910	[iy]	2790
[o]	925	977	1010	[ɔ]	920
[ɯ]	1675	1730	1590	[uw]	950
[ɯ̟]	--	2187	--	[ʌ]	1400

Adapted from Kasuya et al. (1968: 359), Homma (1973: 352) Morimoto (1988: 185) and Peterson and Barney (1952: 183)

We note that for female speakers, F_1 values indicate a similar degree of closeness/height for AE and Japanese high vowels, and a similar degree of openness/low tongue position for [a]. F_2

[35] For space-saving reasons, Homma's fronted [ɯ̟] is placed on the same line as AE [ʌ], but this does not imply equivalence between the two vowels.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

values are also comparable for AE [iy] and Japanese [i], suggesting equivalent anteriority. Compared to the female Japanese [a], its counterpart AE F_2 value suggests a more back articulation, similar to the F_2 of Kasuya et al.'s study, which indicates the backmost of the three Japanese [a]'s. This is a small piece of evidence that that female Japanese [a] may be farther forward than its female AE equivalent. But a more striking interlinguistic difference is found comparing the F_2 values for AE [uw] and those for Japanese [ɯ] and [ɯ̟]. The AE value for [uw] indicates a position of equivalent backness/rounding to the AE mid vowel [ɔ], comparable with those values examined for Japanese [o]. In contrast, the values for [ɯ] suggest a much more anterior position; they are close (although forward of) the F_2 value for the central or backed-central AE [ʌ], while the F_2 value for [ɯ̟] places it nearest horizontally to the AE front vowel [ɛ].

Table 9.6: Comparison of male F_1 values (in Hz), drawn from five studies of Japanese and one of American English.

Ja.	Kasuya et al.	Onishi spectro-gram	Onishi group	Morimoto	Strange et al. 1-mora	Strange et al. 2-morae	AE	Peterson & Barney
[a]	775	750	600	760	618	716	[a]	730
[e]	475	542	450	480	437	443	[ɛ]	530
[i]	263	375	250	310	317	295	[iy]	270
[o]	550	458	450	480	430	423	[ɔ]	570
[ɯ]	363	417	300	380	349	324	[uw]	300

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Adapted from Kasuya et al. (1968: 359), Onishi (1973: 201, 672), Morimoto (1988: 185), Strange (1998) and Peterson and Barney (1952: 183)

Table 9.7: Comparison of male F_2 values (in Hz), drawn from five studies of Japanese and one of American English.

Ja	Kasuya et al.	Onishi spectro-gram	Onishi group	Morimoto	Strange et al. 1-mora	Strange et al. 2-morae	AE	Peterson & Barney
[a]	1163	1375	950	1320	1182	1175	[a]	1090
[e]	1738	1500	1900	1960	1849	1982	[ɛ]	1840
[i]	2263	2208	2220	2360	2077	2243	[iy]	2290
[o]	838	917	750	870	805	730	[ɔ]	840
[ɯ]	1300	1250	900	1320	1171	1160	[uw]	870
--	--	--	--	--	--	--	[ʌ]	1190

Adapted from Kasuya et al. (1968: 359), Onishi (1973: 201, 672), Morimoto (1988: 185), Strange (1998) and Peterson and Barney (1952: 183)

In comparing F_1 values for male speakers across both languages, we note that, like the females', they indicate equivalences in closeness/height for the high vowels and in openness/low tongue position for [a]. Comparison of F_2 values for [a] locate the AE vowel horizontally midway between the backmost Japanese [a] (the Onishi group's) and the backed-central [a]'s of Kasuya et al.'s and of Strange's two data sets. Thus, relative to its male Japanese equivalents, AE [a] is not the farthest back in placement; this differs from the female [a]'s compared and gives a small further piece of evidence for the anteriorizing of female Japanese [a] and the relative backing of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

male Japanese [a]. F_2 values for AE [uw] and for Japanese [ɯ] do agree in one instance, suggesting a comparable back articulation (the Onishi group's [ɯ]).[36] But apart from this, the other five F_2 values for [ɯ] all suggest a placement anterior to AE [uw], and are once again roughly similar to the F_2 value for the central or backed-central AE [ʌ]. We conclude that the more forward position of Japanese [ɯ] relative to AE [uw] appears to be the same for both sexes, unaffected by a male tendency towards the backing of non-front vowels.

In sum, acoustic data indicate that the vertical specifications of Japanese [a], [i] and [ɯ] do not differ very much from their AE equivalents. However, along the horizontal dimension there is a definite difference between the Japanese and AE non-front vowels. The Japanese vowels [a] and [ɯ] are central or even farther forward, and both appear to be articulated with much the same degree of tongue advancement, while their American equivalents are respectively backed-central and back. The horizontal distribution of AE and Japanese vowels is different: while AE [uw] and [ɔ] are articulated farthest back, Japanese [o] appears to be the one true back vowel out of the five. A further contrast with AE may perhaps be seen in the influence of gender on the horizontal placement of [a] for Japanese speakers, but there is insufficient evidence to make a solid claim here.

Locus. As observed at the beginning of this section, no one study of Japanese vowels gives values for F_0 through F_3. Therefore, an approach that stitches together data from a number of sources is the only possible way of ascertaining the presence or absence of locii of main

[36] We note that the Onishi group's [o] and Strange's two-morae [o] have even lower F_2's than those of the AE back vowels (although values fall within the 175-Hz limit of similarity)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

concentration of acoustic energy. Naturally, this can be no more than a partial attempt at the facts.

The relation of F_0 to F_1 values must be surveyed to determine whether or not low locii for Japanese [ɯ] and [i] (analogous to those for AE [uw] and [iy] and for Korean [u] and [i]) are present. Values of F_0 together with F_1 and F_2 have been located only in Homma (1973), presented below.[37]

Table 9.8 F_0 through F_2 values for the point vowels (in Hz) for one female Japanese speaker

	F_0	F_1	F_2
[a]	328	1046	2075
[i]	**350**	**354**	2886
[ɯ]	**342**	**375**	1730

Adapted from Homma (1973: 352, 356)

Bolding on the F_0 and F_1 values for [i] and [ɯ] indicates low locii of acoustic energy concentration for these vowels; values are almost identical for [i] and very close for [ɯ]. We also note here that the separation between the values for F_1 and F_2 indicates the absence of a middle locus for [a]. Since F_3 values are not provided, we cannot determine whether there is a higher locus for Homma's [i].

To discover whether or not this higher locus exists (similar to those found for AE [iy] and Korean [i]), the relations of F_2 to F_3 values from three sources (giving F_1 through F_3) have been surveyed and are presented below; these are male and female values from Kasuya et al. (1968),

[37] But the reader is cautioned (for this and other acoustic data of which Homma is the source) that, however thorough her study, she is reporting just on a single informant.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(presumed) male values measured from the spectrogram in Onishi (1981) and sets from the data of Strange (1998) for one-mora and two-morae vowels.

Table 9.9: F_1 through F_3 values (Hz) for the point vowels of between four and seven adult female and male Japanese speakers

	Female			*Male*		
	F_1	F_2	F_3	F_1	F_2	F_3
[a]	888	1363	3050	**775**	**1163**	2713
[i]	325	**2725**	**3475**	263	**2263**	**3000**
[ɯ]	375	1675	2688	363	1300	2350

Adapted from Kasuya, Suzuki and Kido (1968: 359)

Table 9.10: F_1 through F_3 values (Hz) for the point vowels of one (presumed male) Japanese speaker

	F_1	F_2	F_3
[a]	750	1375	2542
[i]	375	**2208**	**3125**
[ɯ]	417	1250	2417

Measured from spectrograph, Onishi (1981: 201)

Table 9.11: F_1 through F_3 values (Hz) for the point vowels of four male Japanese speakers

	1-mora			*2-morae*		
	F_1	F_2	F_3	F_1	F_2	F_3
[a]	618	1182	2289	716	1175	2348
[i]	317	**2077**	**3027**	295	**2243**	**3123**
[ɯ]	349	1171	2314	324	1160	2309

Adapted from Strange et al. (1998)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In all cases, bolding on F_2 and F_3 values for [i] indicates a high locus of acoustic energy for this vowel.[38] Lacking the F_0 to F_1 range, the existence of lower locii for [i] and [ɯ] cannot simultaneously be ascertained, but their low F_1 values strongly indicate that lower locii are present. Bolding on Kasuya et al.'s F_1 and F_2 values for [a] indicates a middle range locus for the male vowel, but we note its absence for the female [a] and for the [a]'s of Onishi and Strange.

Since the relation of F_1 to F_2 values determines the existence of a middle locus of acoustic energy for [a], (as seen in AE, but not in Korean),[39] data covering this acoustic area are available from six sources. In addition to the four shown above, two more studies, shown below, provide values for F_1 and F_2; these are Onishi (1981, group) and Morimoto (1988).

Table 9.12: F_1 and F_2 values (Hz) for the point vowels of an unspecified number of (presumed male) Japanese speakers

	F_1	F_2
[a]	**600**	**950**
[i]	250	2220
[ɯ]	300	900

Adapted from Onishi (1981: 672)

[38] The reader is reminded that Hamano's (1986) phonosymbolic characterizations of Japanese [i] (quoted above, in Section 9.1.1) appear to refer to the presence of prominent high-frequency acoustic concentration.

[39] For details, see Chapter 8, Section 8.1.2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9.13: F_1 and F_2 values (Hz) for the point vowels of five male and five female Japanese speakers

	Female		*Male*	
	F_1	F_2	F_1	F_2
[a]	990	1600	760	1320
[i]	320	2910	310	2360
[ɯ]	450	1590	380	1320

Adapted from Morimoto (1988:185)

Values (bolded) for Onishi's male group [a] indicate a middle locus, while those for Morimoto's male and female groups do not. Thus, out of the six studies, a middle locus for [a] is found in only two, and both are the vowels of male speakers. We have already observed that the horizontal location of Japanese [a] is variable, i.e., that its F_2 may vary considerably from study to study; moreover, there have been indications that female [a] may be articulated more anteriorly than its male equivalent. We also note that the farther forward that [a] is articulated, the higher will be its F_2, and the consequent greater separation in value from F_1. Thus, fronted [a]'s (possibly tending to be female) will resemble AE [ʌ] or Korean [a] in showing a diffusion of acoustic energy rather than a concentration in the middle range; on the other hand, more backed [a]'s (possibly tending to be male) will show a middle locus, analogous to the AE vowel.

In sum, a survey of formant values from various studies indicates that for the Japanese high vowels we may expect locii of concentration of acoustic energy similar to those for American English, while the horizontal variability of Japanese [a] means that it may or may not possess a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

locus. But, as remarked at the beginning of this section, these observations are based on a rather unsatisfactory patchwork approach; thus, they cannot be considered definitive.

Intrinsic duration. Data here are from three sources: Han (1961), Homma (1973), and Strange et al (1998). Han does not report a difference between one-mora and two-morae durational rankings.

Table 9.15: Ratio of relative durational values of point vowels of three Japanese speakers (sex unspecified)

	V_1
[a]	1.44
[i]	1.17
[ɯ]	1.00

Adapted from Han (1961: 17)

Han's values indicate that [a] is 44%, and [i] 17%, longer than [ɯ].Her greatest-to-least ordering agrees with one of the two sets of values given by Homma's study.

Table 9.16: Values of average point vowel durations (milliseconds)[40] in disyllables for one female Japanese speaker

	Token #	V_1	Token #	V_2
[a]	63	96	39	124
[i]	20	81	49	109
[ɯ]	26	76	28	113

Adapted from Homma (1973: 359)

[40] The disyllables measured cover all four types of pitch-accent found in Homma's own, Kyoto, dialect, but these have been pooled, since she finds that "accent and vowel duration have no significant relationship in Japanese" (Homma 1973: 360). The uneven numbers of tokens have been included so that the reader may be aware of possible bias in these results. All vowels are presumably one-mora, although not specified as such.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

We note that V_2 vowels are always longer than V_1; in either position, [a] has the greatest duration. However, position reverses the high vowels' durational rankings, with the V_1 placement yielding the lower value for [ɯ]. In Strange's data on the durations of one-mora and two-morae vowels, we see a similar phenomenon. Regardless of mora-length, [a] has consistently the greatest duration; of the two-morae vowels, [i] has the least duration, whereas of the one-mora vowels, [ɯ] has the least.

Table 9.17: Values of average durations of one-mora and two-morae point vowels (ms) of four male Japanese speakers

	One-mora	two-morae
[a]	52.79	165.03
[i]	50.53	137.56
[ɯ]	46.70	146.05

Adapted from Strange et al. (1998)

Thus, assuming a correlation between phonosymbolic size and phonetic duration,[41] we would expect <a> to be evaluated as the largest/least-small vowel, whatever mora-length was used for pronouncing the stimulus vowels. We cannot make similar predictions about the other three stimuli; furthermore, Strange et al's (1998) data raise two important questions for this experiment. Firstly, given the absence of recordings made at the time, is there any way of determining whether the Japanese subjects were pronouncing the stimuli with one-mora or two-morae vowels? Secondly, since two of the written stimulus cues (*hee* and *hoo*) were spelt with double vowel letters and two (*hah* and *huh*) with single vowel letters, is it possible that <i> and <u> were pronounced with two-morae vowels, and <a> and <ʌ> with one-mora vowels? These

[41] For details of this issue, see Chapters 2 (the discussion of Sapir (1929: 243) and Newman (1933: 68)) and 7, Section 7.2.3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

questions cannot be definitively answered until presentation of the data, specifically findings on the size dimension (see below, Section 9.2.3); however, predictions can be made here which can then be matched with actual results.

In tentative answer to the first question, an evaluation of <i> as least-large/smallest would argue for a two-morae pronunciation length, while evaluating<ʌ> or <u>[42] as least-large/smallest suggests a pronunciation length of one mora. Answering the second question, if <a> and <ʌ> are found to be judged as nonlarge/small and if there is an <ʌ/a> collapse, while if <i> and <u> are judged as large/nonsmall and collapsed, then we have evidence that the stimulus spelling influenced mora-differentiation between the vowels produced. If, on the other hand, the stimuli are evaluated differently, or different pairs are collapsed (such as <ʌ/i> or <u/a>), then evidence will indicate a uniform pronunciation length for all four stimulus vowels.

Intrinsic amplitude. Here we have data from Homma only.

Table 9.18: Values of average point vowel amplitudes (dB) for one female Japanese speaker

	token #	amptde
[a]	90	72
[i]	71	58
[ɯ]	51	66

Adapted from Homma (1973: 358)

Ranking the point vowels, we note that [a] has the greatest amplitude, followed by [ɯ] and then [i]. Thus, if amplitude has a phonosymbolic effect, we would predict for the Japanese data that

[42] This is based on the hypothesis that both derive from /ɯ/, argued in Section 9.1.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

there would be an opposition between the (possibly collapsed) non-front vowels <a>, <u> and <ʌ>, perceived as louder, and <i>, as softer.

Intrinsic F_0. Data here are taken from Homma (1973) and Nishinuma (quoted in Whalen and Levitt, 1995). Homma gives the following average values (in Hz.) for the point vowels examined in her study, ordered from greatest to least:

Table 9.19 : F_0 values (Hz) for one female Japanese speaker's point vowels

[i]	350
[ɯ]	342
[a]	328

Adapted from Homma (1973: 356)

This ordering agrees with those of Korean and English in giving [a] the lowest F_0, but differs from them in reversing the rankings of the high vowels, since AE [uw] and Korean [u] have the highest F_0's. On the other hand, Nishinuma, with a sample of five male informants, quoted in Whalen and Levitt (1995),[43] gives an ordering similar to the Korean and English:

Table 9.20 : F_0 values (in Hz) for five male Japanese speakers' point vowels

[i]	142
[ɯ]	148
[a]	136

Nishinuma, adapted from Whalen and Levitt (1995: 352)

---

[43] This is a survey, comparing the intrinsic F_0's of [u], [i] and [a] in 31 languages, and covering 11 out of the 29 language families.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Homma and Nishinuma are the only two references are given by Whalen and Levitt for Japanese F_0; thus, it follows that a definitive intrinsic F_0 ranking cannot be established for the high vowels. Still, however [i] and [ɯ] are ranked, [a] contrasts with them in having the lowest F_0,[44] and the fundamental-based contrast of [a] to [i]/[ɯ] may play a role in phonosymbolic evaluations. If the Japanese data show an opposition in evaluations between <a> and the (possibly collapsed) triad <i>-<u>-<ʌ>,[45] then we may have evidence that F_0 is at work.

[44] Whalen and Levitt's study indicates that this is true for every language surveyed, and they conclude that it is universal and intrinsic, "a direct result of vowel articulation" (Whalen and Levitt, 1995: 363).

[45] See again the argument in section 9.1.1 for allophones of /ɯ/ as underlying pronunciations of <u> and <ʌ>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

9.2 Analysis of the Japanese data

Figure 9.1 below presents in summary form the stimulus rankings on the six semantic scales, derived from the responses of the Japanese speakers. It corresponds to Figures, 7.2. and 8.2, in Chapters 7 and 8, which summarize, respectively, the responses of the AE and Korean subjects. Like the earlier figures, it is based on indices and is accurate in scale; rankings are integrated with consensus data, with only the more unusual cases of non-differentiation being indicated.

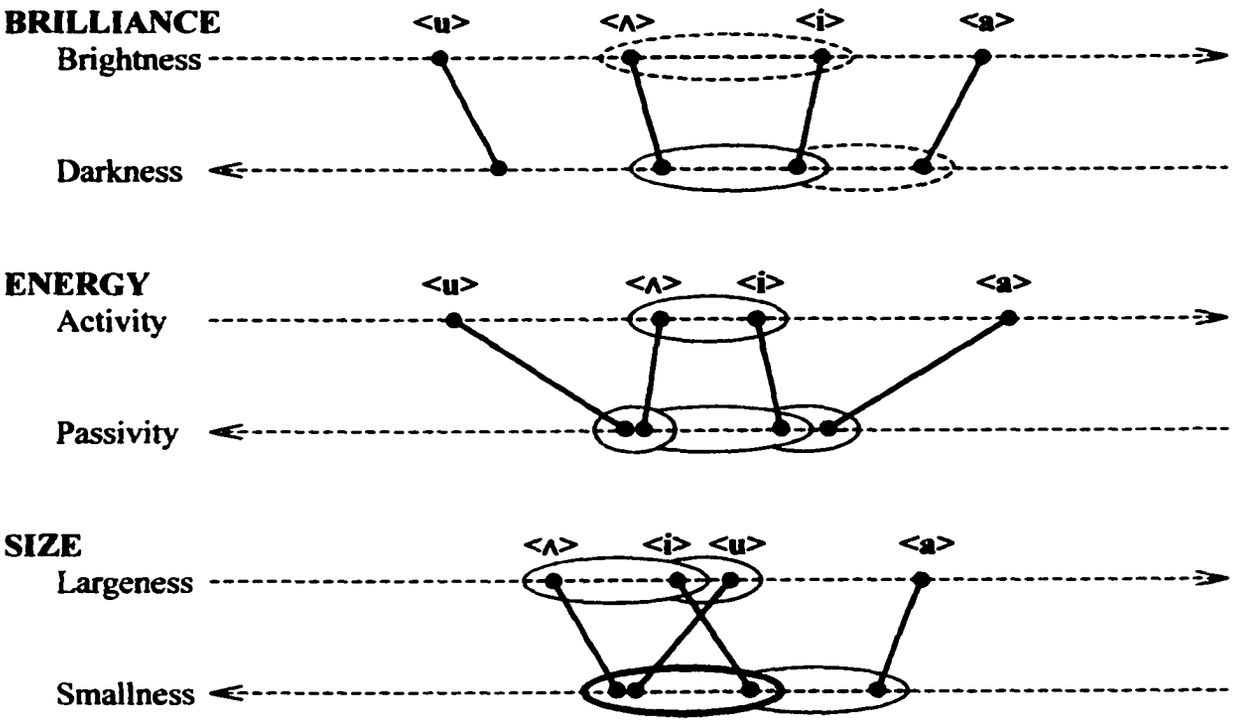


Figure 9.1, Data patterns (Japanese speakers). Ranking index values and consensus data for salient and non-salient sub-scales of three semantic dimensions. Ellipses surround stimulus pairs not differing by statistical test, $p>.05$; dashed outlines indicate instances of trend, $.10>p>.05$; bolding indicates a special case with no significant contrasts among a set of *three* stimuli.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A first point, on inspecting Figure 9.1, is the broad similarity between the sets of stimulus rankings obtained on the brilliance and energy dimensions; as with AE and Korean data, all of the relevant scales pattern together. The Japanese rank order of stimuli agrees with the Korean but disagrees with the AE arrangement in the ranking of <a>, and not <i>, as brightest/least dark and most active/least passive. At the opposite endpoint of the scales, the Japanese disagree with both of the other groups in ranking <u>, rather than <ʌ>, least bright/darkest and least active/most passive.

Also, as with the AE and Korean findings, size displays a different stimulus ranking from those of the other two dimensions. But whereas the AE and Korean size rankings and configurations agree with each other, those of the Japanese agree with them only in one aspect: the ordering of <a> as largest/least small.[46] While AE and Korean collapse <u/a> on both salient and non-salient size scales, Japanese places <a> on a separate rank step on largeness. Also, whereas <i> is least large/smallest for the AE and Korean subjects, the Japanese evaluate it as non-small, collapsing it with <a> on smallness. Japanese ranks <ʌ> as least large/smallest, and the collapsed <i/u> fall in a middle position. Given that <ʌ>, rather than <i>, is ranked least large/smallest, this ordering provides some indication that one-mora length was used for stimulus pronunciation, since it agrees with Strange's (1998) one-mora durational rankings for [i] and [ɯ].[47] However, though <ʌ> occupies an endpoint, it is also collapsed with <i>; thus, the rankings of these two stimuli

[46] We except here the asymmetrical Korean smallness ranking of <u> as least small.

[47] Strange et al.'s (1998) durational data for [i] and [ɯ] indicated that of the two-morae versions of the vowels [i] was shortest, while of the one-mora versions, [ɯ] was shortest. For details, see Section 9.1.2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

should be counted as an indication, not evidence, of mora-length: they lack a statistical basis and may be the product of chance.[48]

Following our examination of the AE and Korean data, the Japanese ranking similarities between the brilliance and energy scales are the third interdimensional agreement of this type to be found. This is additional evidence that, regardless of language group, subjects were accessing a semantic metadimension of which brilliance and energy were felt as subdimensions. The fact that brilliance/energy rankings are *not* agreed-upon interlinguistically, even though intralinguistic similarities between dimension results are clear, further strengthens the metadimension theory and supports categorization of the six scales into a brilliance/energy grouping of four (the 'vitality' scales) and a size grouping of two.

A second point arising from inspection of Figure 9.1 is the symmetricality of rankings for the salient and non-salient scales, with one minor exception.[49] Also, as in the previous two cases, the non-salient scales have narrower spacing than the salient, but spacing/range disparities are considerably less for brilliance and size than they are for energy. Here, the agreement of the Japanese data with those of Korean and AE groups is striking, as the following table illustrates.

[48] For discussion of this point, see Chapter 6, Section 6.2 .

[49] The exception arises with <i> and <u> on the size scales.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9.21: American English, Korean and Japanese response range and discrepancy values for salient and non-salient scales of three semantic dimensions

	AE			Korean			Japanese		
Dim.	*sal.*	*disc.*	*n-sal.*	*sal.*	*disc.*	*n-sal.*	*sal.*	*disc.*	*n-sal.*
Brill.	.710	*.096*	.614	.591	*.029*	.562	.530	*.115*	.415
En.	.612	**.165**	.447	.541	**.198**	.343	.541	**.343**	.198
Size	.342	*.059*	.283	.436	*.044*	.392	.359	*.106*	.253

Bolded numbers indicate that indicate that in all three data sets, the energy dimension shows the largest response range discrepancy. Underlined italicized numbers indicate that in each group, brilliance and size have similarly-valued, lesser, range disparities (the two values diverge more for the AE than the Asian language speakers). Since this tendency to range-shrinkage on the passivity scale includes the English speakers, we have an indication that possible difficulties in making passivity evaluations were not simply generated by problems in translating the adjectival term "passive." But we also note that the Japanese energy-range disparity value is almost double that of the Korean speakers'.[50] The question is, therefore, whether two factors may have not been in play, one interlinguistic, arising from ease-of-judgment differences between attributing activeness and passiveness, and the other intralinguistic, caused by translation difficulties. Accordingly, Figure 9.2 below presents the totals of translations (correct or not) of adjectival terms in the Japanese responses to the fourth page of the questionnaire, and compares them with the numbers of mistranslations.[51]

---

[50] We note that all three discrepancy values are higher for the Japanese group than for the other two groups, suggesting that they may generally have found the non-salient terms more difficult to apply than the salient.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

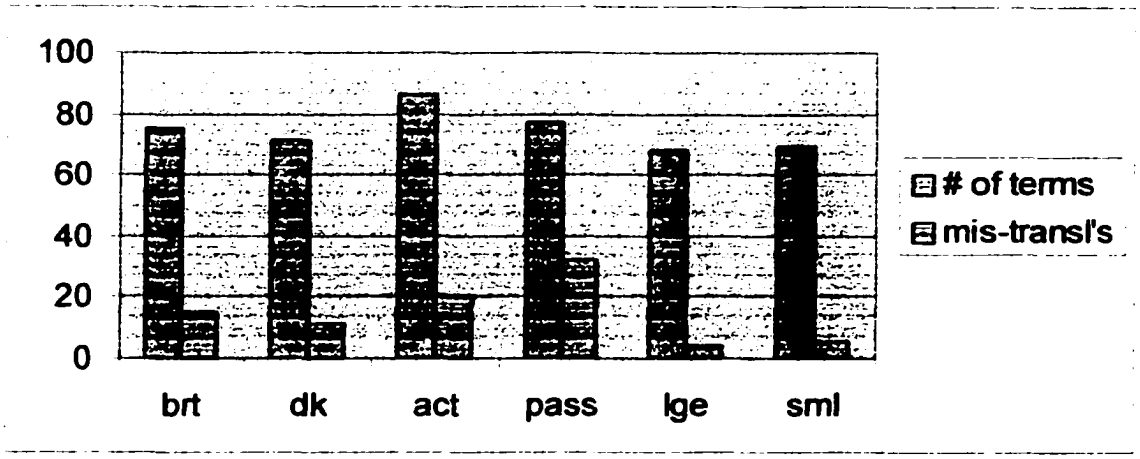


Figure 9.2: Totals of Japanese translations of adjectival terms, compared with numbers of mistranslations.

Fiftynine subjects responded to the fourth page of the questionnaire, but some gave more than one translation for a term; the total number of translations ranged between 86 for *active* and 69 for *small* As Figure 9.2. indicates, salient and non-salient size terms yielded similarly low proportions of mistranslations: 6% (4/68) for *large* and 7% (5/69) for *small*. The proportions of mistranslations of the salient and non-salient brilliance terms were dissimilar, but in both cases greater than those for size: 20% (15/75) for *bright*, and 15% (11/71) for *dark*. Of the energy terms, *active* was comparable with *bright*, producing 23% (20/86) mistranslations, but *passive* gave by far the highest proportion of mistranslations: 42% (32/77)—almost double that of the salient term, and six times that of the size terms. Thus, results indicate that the translation of *passive* appears to have been especially problematic for the Japanese speakers. Nor can this be

[51] Page 4 of the questionnaire requested the ESL subjects' native language translations of the semantic scale adjectival terms used. (For details of this page, see Chapter 6, Section 6.2.) Ms. Takako Katsuno ably provided back-translations from the Japanese.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ascribed to their inexperience with English, since, as mentioned in Section 9.1.1, ninety percent had been studying English for over two years. [52]

To sum up, two factors, semantic dimension and salience, appear to have affected response-range width. If the width of a scale's response range can be taken as indicating ease/difficulty of phonosymbolically judging stimuli on it, then Japanese subjects apparently found that activity and brightness were almost equally easy to use as phonosymbolic measures. In the order of the descending value of ranges, darkness appears to have been somewhat more difficult than either; more difficult still seem to have been largeness followed by smallness, with passivity by far the most difficult. However, if the complexity of configuration of stimuli on a scale is also an indicator of difficulty, then smallness appears to have been equally—or even more—problematic.

A third point raised by inspection of Figure 9.1 is, once again, the importance of stimulus contrasts arising from the pairing of stimuli in the test items. Here, we see a wide range of differentiability among the pairs, with dimension, stimulus partner and salience all as possible influencing factors. Two pairs, <ʌ/a> and <a/u>, yielded crisply-differentiated evaluations, showing consensus across all six scales. An interaction between dimension and pairing is shown by <i/u>, with consensus for all except the size scales. The pair < ʌ/u> is evaluated less well,

---

[52] I am informed (by Ms. Takako Katsuno, personal communication) that what may have made the term *passive* confusing for Japanese speakers—and hence more problematic to translate—is a lexical-cultural factor. Both strongly positive and strongly negative synonyms exist for it in the language: on one hand "philosophical, "calm," "contemplative," and on the other, "wimpy," "victimizable". (In contrast, for Americans, positive synonyms and connotations of *passive* would be harder to find outside of highly specialized contexts, because less sanctioned by culture and religion.)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with consensus on both brilliance scales, but only on the salient energy and size scales, while <i/a> appears yet more sensitive to salience, showing consensus only on the salient scales of all three dimensions. Finally, the pair <ʌ/i> were overall poorly differentiated, yielding a cluster of non-consensus on all six scales.

The importance of clusters. So far, collapses have been comparatively rare in the overall picture (occurring in 19.5% of the AE, 30.6% of the Korean and 36.2% of the Japanese responses). Clusters of collapses are much more rare: there are none in the AE data and only one each in the data of the Korean and Japanese groups. This strongly suggests that such cases are not simply the products of matching one, or even two, phonosymbolically irrelevant scales to a stimulus pair. Instead, regardless of dimension or salience, clusters seem to demonstrate a consistently confused response to requests for judgements about two stimuli. Our argument is strengthened by the particular rarity of collapses on salient scales (in the total of each group's collapses, found in 2/7 (28.6%) for AE, 4/11 (36.4%) for Korean and 4/13 (30.8%) for Japanese). In both Korean and Japanese data, three of those four salient scale collapses are found in the clusters. [53]

The pronunciation of <ʌ> and <u>. The appearance of the Japanese <ʌ/i> cluster, together with other stimulus contrast data, gives evidence about which of their first-language vowels Japanese subjects used or referred to in their renditions of <ʌ> and <u>. The reader will recall that two L1 candidates, [a] and [ɯ], were suggested as possible bases for their pronunciation of <ʌ>.[54] It was

[53] We have already noted a tendency in the two data sets previously examined for the non-salient scales *passive, small* and *dark* to evoke less robustness of response and hence less cleanly separated stimulus scores, than the salient scales *active, large* and *bright*. Figure 9.1. strongly suggests that this tendency is also shared by the Japanese data.

[54] For the full discussion of 'the problem of <ʌ>,' see above, Section 9.1.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

hypothesized that if subjects chose [a], then the pair <ʌ/a> might show a particular tendency to collapse; instead, results show <ʌ/a> as perfectly differentiated, and thus [a] to have been an unlikely choice. If, on the other hand, [ɯ] had been chosen, then <ʌ/u> would seem especially liable to collapse. Instead, results indicate that this pair shows fair differentiation (on 4/6 scales, and without salient scale collapses), suggesting that the subjects had found some way of phonosymbolically and articulatorily separating its member vowels.

However, if, as outlined in Section 9.1.1, the fronted allophone of /ɯ/, [ʉ], had been chosen for <ʌ>, and the backed, rounded allophone [ɯ^{w}] for <u>, we have an explanation of why <ʌ/u> could be satisfactorily differentiated. Also, a further consequence of the front articulation of [ʉ] as <ʌ> could well have been a tendency towards indifferentiability from <i>,[55] and this is indeed what we find in the clustering of <ʌ/i> collapses. Thus, stimulus contrast data provide solid evidence that the first-language referent for <ʌ> was [ʉ], and for <u> was [ɯ^{w}]. Additional evidence that this clustering could have been caused by the front articulation of [ʉ] is that the [h]'s of the four stimuli are likely to have shown assimilatory variation (Akamatsu 1997) in those subjects with stronger L1 interference. This would have resulted not only in similarities between the vowels of <ʌ> and <i>, but also in their initial consonants, and in contrasts of both vowels and initial consonants with those of the other two stimuli.[56]

---

[55] The reader is reminded of Homma's (1973) observation that her [i] and her fronted version of [ɯ] overlapped each other. For further details, see Section 9.1.1.

[56] Thus, both the fronted [ʉ]and [i] would have triggered [ç] the palatalized allophone of /h/, resulting in similar initial consonants for both stimuli: [çi] and [çʉ] for <i> and <ʌ>. Assimilatory variation of /h/ would have also produced two different allophones for the initial consonants of the other two stimuli: [ha] for <a> [Φɯ^{w}] for <u>. See also Akamatsu (1997:87).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Stimulus spellings and mora-length. Stimulus contrast data found in Figure 9.1 also answer the question of the possible influence of *hVh* vs. *hVV* stimulus spelling on the duration of pronunciations, and thus on size evaluations.[57] It was argued that if orthography had been an influence, <ʌ> and <a> (those stimuli spelled with a single vowel letter) should have been considered as shorter (one-mora) and thus evaluated as non-large/small. Furthermore, when paired, they might show an overall tendency to collapse. In fact, as we have seen, <ʌ/a> are always differentiated and are found at opposite extremes of both size scales. It would therefore appear that the *hVh* spelling did not have an effect in associating the two stimuli spelled this way, and did not influence size evaluations.

It was also argued that spelling would have been indicated as an influence if <i> and <u> (those stimuli spelled with a double vowel letter) had shown a similar overall tendency to collapse when paired, and if they had been ranked as large/non-small (i.e., considered as two-morae vowels). At first glance, it seems that the *hVV* spelling may have had some effect, since we see an <i/u> collapse on both size scales (although nowhere else). But these stimuli are evaluated in the middle-to-lower range in terms of magnitude; their non-large status is most clearly seen on the largeness scale, where both are differentiated from the largest stimulus, <a>, and form a chain with the least-large <ʌ>. Our conclusion must be that single vs. double vowel spellings probably had no effect on the mora-length of pronounced stimuli, or on stimulus evaluations.

[57] For details of this discussion, and for references establishing the connection between phonetic duration and phonosymbolic size, see Section 9.1.2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The data will now be treated in detail, following the scheme found in Chapters 7 and 8. Subsection 9.2.1 will discuss brilliance, 9.2.2 energy, and 9.2.3 size. Section 9.2.4 summarizes findings and presents a semantic profile of each of the stimulus vowels as evaluated by the Japanese subjects.

9.2.1 Brilliance dimension

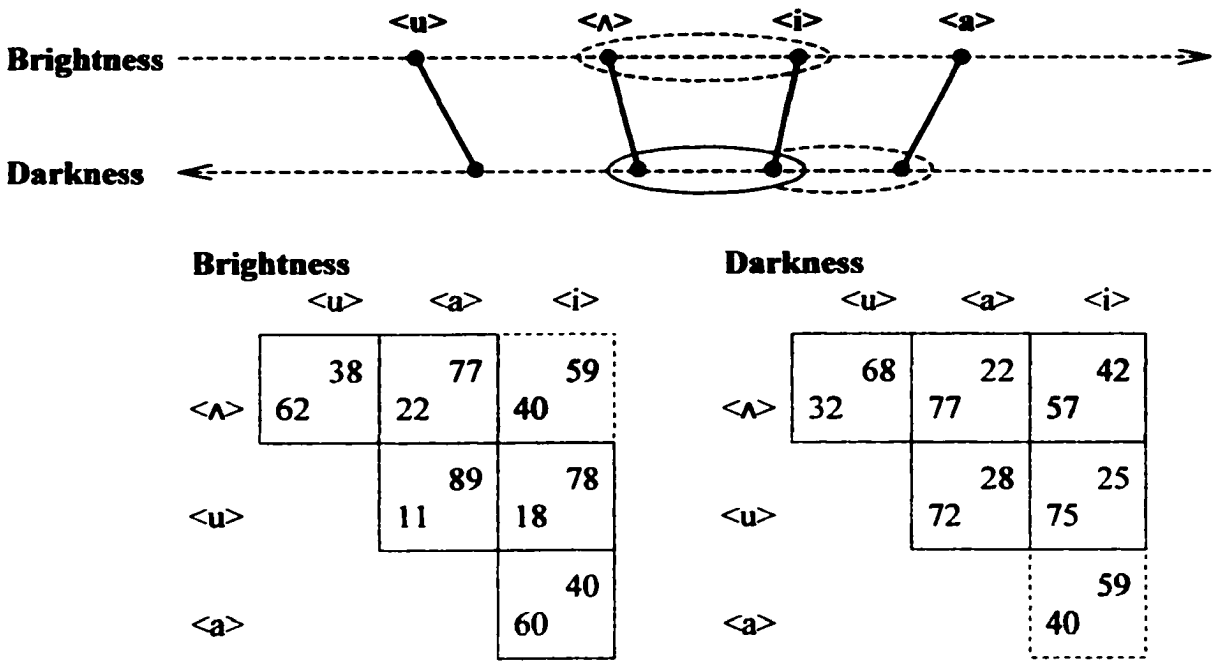


Figure 9.3: Brilliance dimension data arrays (Japanese speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing significantly, $p>.05$; dotted lines indicate trend, $.10>p>.05$.

Following the form of the two preceding chapters, Figure 9.3. above extracts the brilliance dimension data previously presented as part of Figure 9.1's overall summary of the data for Japanese speakers and combines it with the frequency matrices from which it derives, permitting a detailed consideration of subjects' evaluations of the stimuli on brightness and darkness.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

General remarks on the scales. The upper panel of Figure 9.3 shows that on both scales rankings are collapsed into three steps, with <ʌ/i> found in the middle position. However, we see differences in spacing configurations between the scales: on the non-salient scale there is a further collapse of the lower two steps into a chain, with <i> indeterminate between least-dark <a> and second-darkest <ʌ>. Also, comparing index value ranges, we find a definite shrinkage (of .115 points: see also Table 9.21 above) for the non-salient scale, and, in addition, the lower panel of 9.3 indicates slightly lesser robustness of response for the non-salient scale. We see this most clearly in the case of <u/a; for <u/i> and <ʌ/i> we find modest differences, and a questionnable one for <a/i>. There are no difference in scores on either scale for <ʌ/a> and a modest difference in favor of the non-salient scale for <ʌ/u>.

Interpretation of stimulus rankings. Japanese disagrees with English, but agrees with Korean in the ranking of <a> as brightest/least-dark stimulus and in the collapsing of <a/i> on darkness. However, Japanese differs from Korean in placing <a> and <i> on separate rank steps on brightness. In explanation of this extreme ranking for <a>, two of the arguments offered in the preceding chapter are relevant.

Firstly, although it lacks [i]'s high locus of concentration of high frequency acoustic energy, [a] possesses other features (high acoustic amplitude and wide jaw opening) capable of triggering a brightness judgment. These characteristics were argued to constitute the linked acoustic-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

articulatory 'amplitude' feature, a competitor in vitality evaluations with the 'locus' feature.[58] As a non language-specific quality of [a], the amplitude feature was argued possibly to have influenced the AE brilliance rankings for <a> (as second-brightest/second-least-dark). But in the Japanese case, if our hypothesis about <ʌ>'s being pronounced by these subjects with the vowel [ɯ̟] is correct, then <a> was also the only open stimulus vowel for these speakers—i.e., the only one with the amplitude feature. Thus, a language-specific contrast would have reinforced the opposition between <a> and the other three stimuli.

Secondly, several of the studies consulted (Isshiki, 1965, Homma 1973, Vance 1987, Akamatsu 1997) suggest that Japanese [a] can have a front-central, even a front, tongue placement, parallelling the Korean case. Shared non-backness would associate [a] with [i], which could explain the <a/i> collapse on darkness; non-backness may also account for the ranking of <i> as second-brightest on brightness.

At the other extreme of the brilliance scales, we see a (so far) unique ranking for <u>. On both scales <u> is differentiated from its rank neighbor; unusually, the placement is even more emphatic on the non-salient scale, where <u> is the only independently-ranked stimulus, resulting in a configuration of one dark, vs. three non-dark stimuli. To explain this, we must look for a unique feature, or features, of <u>, and one(s) which could serve as better darkness triggers than the feature of frequency diffusion (muddiness)[59] which has been hypothesized as giving <ʌ> its darkest status in the AE and Korean data.

---

[58] For discussion of the locus feature, see Chapter 7, Sections 7.2.1 and 7.2.2 , and for discussion of the amplitude feature, see Chapter 8, Section 8.2.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Just as <a> has been argued as the only stimulus to have the amplitude feature, so <u> has been argued as the only stimulus to have the linked 'long-tube' features of backing and rounding.[60] These place it in horizontal articulatory contrast both to the central (or front) <a> and to the other two stimuli with high vowels: <i> with its front [i] and <ʌ> with its fronted [ɯ̟]. Although acoustic studies of the Japanese [ɯ^{w}] allophone of /ɯ/ could not be located, we would expect its lengthened front cavity to have produced a further lowering of F_2. Thus, for the Japanese subjects, <u> should have been in acoustic contrast with <ʌ>, as well as with <i>. In addition, we can argue for another acoustic contrast, in perceived intensity, to have been produced by <u>'s lip rounding. The stimulus <u> may have seemed markedly softer than the other two high vowel stimuli because rounding reduced the lip opening and thus its amplitude. Striking evidence for this is given in Hamano's (1986) dissertation on phonetic symbolism in Japanese, where we find a description of [ɯ] as evoking expulsion of air "through *a small opening, such as...the protruded lips*" (Hamano 1986: 100, emphasis added). Thus, <u>'s rounding and consequent damping would be in contrast with the neutral lip positions[61] and neutral intensities of <ʌ> and <i>. Although it treats American English vowels, Tolman's (1887, 1904) tone-color vowel scale[62] provides support for our rounding-based interpretation of the Japanese evaluations of <u> as a dark stimulus. The two most backed and rounded AE vowels [uw] and [ɔ] are found at the *dark* endpoint of the scale.

---

[59] For a discussion of the "muddiness" of central vowels, see Masson (1952: 40).

[60] See Sections 9.1.1 and 9.2 for arguments that <u> for these subjects was pronounced with reference to their backed and rounded [ɯ^{w}].

[61] See Vance (1987) and Akamatsu (1997) for the absence of lip spreading on Japanese [i].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Next, we turn to an explanation of the equally unusual mid-range status of <ʌ> (allowing it to be ranked as non-dark in a chain), and of its collapse with <i>. It can be argued that the answer to both questions is found in the hypothesized fronting of <ʌ> (i.e., its L1 referent, [ʉ̟]), together with the absence of lip spreading for <i>. So far, muddiness (absence of locus) has been claimed as responsible for the extreme non-bright/dark evaluations of <ʌ> in the AE and Korean data. But *if the Japanese subjects' [ɯ] was fronted to [ʉ̟] to distinguish it from [ɯ^{w}]* (that is, to keep <ʌ> and <u> separate), then <ʌ>'s F_2 would have risen and its timbre thus have become less muddy as it approached the anteriority of <i>. Furthermore, lip spreading, the labial feature which otherwise would have helped to distinguish <i> from <ʌ> was not present. Finally, if <a> was perceived as the loudest stimulus, because of its openness, and <u> the softest, because of lip rounding, then <i> and <ʌ>, neither open nor damped by a constricting lip gesture, should have shared a middle (i.e., neutral) ranking in terms of perceived loudness.

To sum up, two dominant triggers (groupings of linked phonetic features) appear largely responsible for the brilliance rankings: *amplitude* (jaw openness, plus acoustic intensity), connoting brightness, and *rounding* (lip gesture and tongue backing, plus acoustic damping) connoting darkness. The two triggers are functioning in harmony with each other:

<a>	+ amplitude, − rounding
<i>, <ʌ>	− amplitude, − rounding
<u>	− amplitude, + rounding

Each trigger yields groupings of a single possessor vs. a triad of non-possessors:

---

[62] For a diagram and discussion of the tone-color scale, see Chapters 2 and 7, Section 7.1.3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

+ amplitude	**− amplitude**
(open/loud)	(non-open/non-loud)
<a>	<i>, <u>, <ʌ>
+ rounding	**− rounding**
(back/soft)	(non-back/non-soft)
<u>	<a>, <i>, <ʌ>

Thus, the polarization of <a> and <u>, as respectively brightest/least dark and darkest/least bright is explained by unique, respective possession of amplitude and rounding. In contrast, their shared lack of amplitude and rounding together account for the collapse and middle status of <ʌ/i> on both scales; shared non-backness explains the <a/i> collapse as non-dark, leading to the <a>-<i>-<ʌ> chaining, on darkness.

Prominence. However, there is one aspect of the data which cannot satisfactorily explained by reference to these features. This is the consistent differentiation of <ʌ> from <a>: since these vowels also share non-backness (i.e., non-rounding), they too might be expected to have collapsed, particularly on the non-salient scale, where less differentiated responses are likely. To account for this differentiation, and give additional explanation for the <a/i> collapse, a third, more general set of contrasts, in *prominence*, is here suggested.

We begin with the <a/i> collapse. In addition to shared non-backness/non-roundedness an equivalence[63] may have been operating between two disparate brilliance-triggering features of these vowels: amplitude and high locus. Evidence for the comparative 'prominence,' though for different reasons, of both [a] and [i] compared to [ɯ], is given in Hamano's (1986) accounts of

[63] For a similar argument concening <a/i> in the Korean data, see Chapter 8, Section 8.2.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the phonosymbolic 'personalities' of the point vowels. In auditory terms, [ɯ] is characterized as expressing "a mellow unobtrusive sound;" this contrasts with the "shrill... high-pitched... piercing... tense" sounds expressed by [i] (Hamano 1986: 97-102, 146). A different form of phonosymbolic prominence, also contrasting with [ɯ], is found in the sounds expressed by [a]: "colorful... gaudy... *loud*... explosive... *wide-ranging*... [connoting] conspicuousness" (Hamano 1986: 98-102, emphasis added).

Although she does not analyze the phonetic sources of these symbolisms, Hamano's account of the vowels supports the hypothesis that for <a> prominence is triggered by amplitude, while for <i> the trigger is high locus, plus the acoustic carrying power of the high frequencies, which has been termed *sharpness*.[64] Following her characterizations, we can extract a general feature contrast between pairs of stimuli: the prominent <a> and <i> and the non-prominent <u> and <ʌ> (both being argued to derive from /ɯ/). Thus, the differentiation of <a> from <ʌ> on brilliance (as indeed on all the other scales) can be explained as due to differences in prominence pairing, which cut across the amplitude-rounding scheme, ensuring the separation of these stimuli.

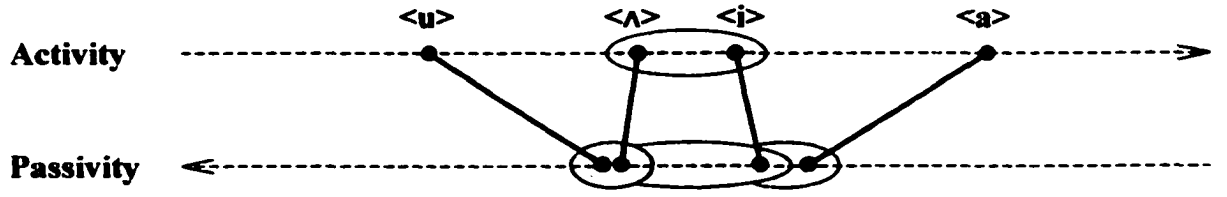
Even though it is claimed that <a>'s amplitude makes it a prominent vowel, and <u>'s rounding/damping make it non-prominent, amplitude and rounding cannot be taken simply as synonyms for prominence and non-prominence. In fact, prominence pairings cut across the triadic groupings proposed for amplitude and rounding. Thus, although the prominent <a> and <i> share non-roundedness/non-backness, prominence cannot be equated with this feature,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

because the third member of the non-rounded triad, <ʌ>, is not prominent. Similarly, though <i> lacks amplitude (which groups it with <u> and <ʌ>), it is differentiated from them by its prominence.[65] While these points appear to make the feature of prominence a complicating factor, it will be seen below, in the discussion of the energy and size dimensions, that it allows us to account for a number of otherwise inexplicable collapses and differentiations.

9.2.2 Energy dimension

Figure 9.4 presents the data arising from judgments on the energy scales by the Japanese speakers.



[64] For a discussion of <i>'s 'sharpness,' see Chapter 7, Section 7.2.1

[65] On prominence grounds, it can also be argued that the brilliance rankings of <ʌ> and <i> may not have been completely due to chance, despite the collapse of these vowels, but instead reflect a felt difference between them. Although argued as phonetically similar to (non-prominent) <ʌ>, <i>'s prominence (i.e., somewhat higher locus and sharpness) may have given it a small 'edge,' moving it slightly further toward the bright/non-dark ends of the brilliance scales and closer to <a>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

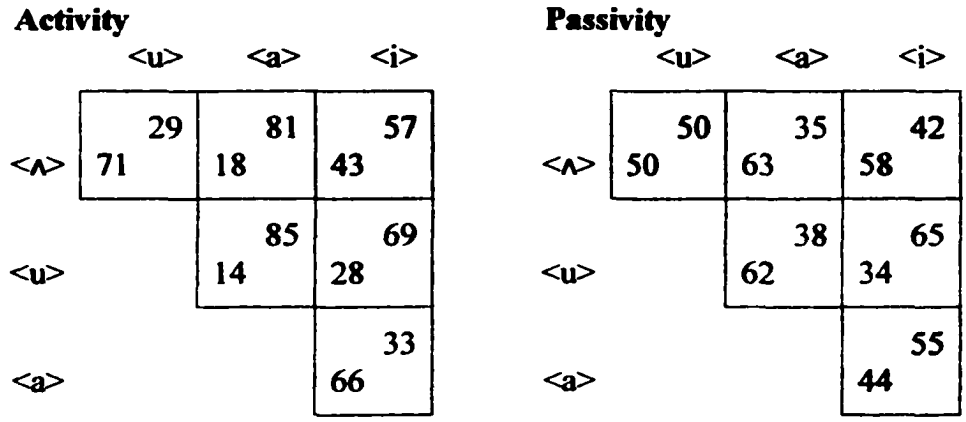


Figure 9.4, Energy dimension data arrays (Japanese speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing significantly, p>.05; dotted lines indicate trend, .10>p>.05.

General remarks on the scales. The upper panel of Figure 9.4 shows that on both scales rankings are collapsed into three steps with <ʌ/i> in the middle position. However, there are striking differences in spacing configurations: on activity we see three separate steps, while on passivity all three steps are collapsed into a chain with two indeterminately placed stimuli. We find <ʌ> indeterminate between most-passive <u> and second-least-passive <i>, and <i> indeterminate between second-most-passive <ʌ> and least-passive <a>. Comparing index-value ranges, we also note a remarkably large shrinkage for the non-salient scale (of .343 points: see also Table 9.21 above). And finally, the lower panel of 9.4 shows a definite diminution in robustness of response on the non-salient scale. This is seen most clearly in the scores of the four pairs <ʌ/u>, <ʌ/a>, <a/u> and <a/i> but is also found for <u/i>. Only for <ʌ/i> do we find a minimal difference in favor of the non-salient scale.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Energy compared to brilliance. The collapse of <ʌ/i> is found on both salient and non-salient scales of both dimensions; we also note that the collapse of <a/i>, and the consequent chaining of <a>-<i>-<ʌ>, occurs on both non-salient scales. Response ranges are similarly wide for the salient scales (.530 for brilliance and .541 for energy), and on both dimensions, range shrinkage is in the non-salient direction. However, the shrinkage is doubled for energy; in addition, we see more extensive chaining on the non-salient energy scale, where <u> uniquely loses its independent ranking. On the non-salient brilliance scale, there is a slight decrease in robustness of response for three out of six stimulus pairs, with a small comparative lessening in majority scores. A decrease in the same direction is found on the non-salient energy scale, but it affects five pairs, and majority scores are more notably diminished. In sum, the salient scales of both dimensions yield comparable results, and both the non-salient scales show similar tendencies towards lack of differentiation. But these are definitely stronger on the non-salient energy than on the non-salient brilliance scale.

Interpretation of stimulus rankings. Although the rankings for salient and non-salient scales are symmetrical, the distinctions are much clearer on the activity scale than they are on passivity. Therefore, we will first focus on activity in accounting for the independent endpoint rankings of <a> and <u>. Then the middle placement on both scales of the collapsed <ʌ/i> will be treated, and finally the 'extras' of the passivity scale: the additional collapses of <a/i> and <ʌ/u>, and the intermediate placements of <ʌ> and <i>.

The points made about brilliance triggers at the close of Section 9.2.1 are applicable here; these can now be broadened to be called 'vitality triggers'. Thus, the amplitude feature (acoustic

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

intensity and jaw openness[66]) can be seen as generating the independent ranking of <a> as most active/least passive, just as it was argued to have ranked <a> as brightest/least dark. We can claim an even better phonosymbolic fit between energy and amplitude than between brilliance and amplitude, due to the direct iconic connection to be made between perceived intensity (loudness vs. softness) and activeness vs. passiveness. In addition, for the Japanese subjects, like the Koreans, amplitude may have been interpreted as a cultural "areal feature," of self-assertion;[67] the connection between self-assertiveness and activeness is clear.

Similarly, the rounding feature unique to <u>, appealed to in explanation of its brilliance rankings, will account for its placement on the energy scales: <u> is opposed to the other three stimuli via its lip gesture, which implies also backing[68] and acoustic damping.[69] Thus <u> would have been the vowel with lowest intensity out of the three vowels lacking the amplitude feature, <i>, <u> and <ʌ>. As in the case of <a>, perceived intensity can be seen as particularly important in triggering <u>'s endpoint energy rankings, due to the iconic connection to be made between auditory softness and passivity. Tolman's (1887, 1904) tone-color vowel scale, although not

---

[66] See the discussion above in Section 9.2.1, where it is claimed that <a> was the only open stimulus vowel for these subjects because the allophones of the close/high phoneme /ɯ/, [ɯ̟] and [ɯ᷄], are considered respectively to have been first-language models for the stimuli <ʌ> and <u>. For further details, see also the articulatory and acoustic descriptions of Japanese [a] in Sections 9.1.1 and 9.1.2.

[67] For a discussion of the phonosymbolic equation between jaw-openness and assertiveness in Korean, see Chapter 8, Section 8.2.2. Further support for this interpretation is also to be found in Hamano's commentary on the use of [a] in Japanese expressives as indicating "that the event, being colorful or gaudy, *stands out*" (Hamano 1986: 98, my emphasis).

[68] For evidence suggesting that Japanese [a] is non-back (central, even anterior) and therefore can be grouped with the non-back [i] and [ɯ], see Homma (1973), Akamatsu (1997) and discussions of centralizing/anteriorizing tendencies of Japanese [a] in Sections 9.1.1 and 9.1.2

[69] It is expected that <i> was pronounced without lip spreading, so that it had the same neutral lip placement as <ʌ>, and therefore <u> was the only stimulus vowel with a constricting (i.e., a damping) lip gesture. See Section 9.2.1 for details of these arguments.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

specifically including *active/passive* amongst its attributes, appears once again of relevance to the evaluations of <u>. The two most backed and rounded AE vowels [uw] and [ɔ] are ranked at the *sad-slow* endpoint of the scale, indicating possession of qualities close in connotations to *passive*.

Finally, the more global feature of prominence can be seen as an additional vitality trigger. It is easy to trace a relation between vocalic prominence (<a>'s amplitude, <i>'s sharpness), the general quality of noticeability, and brightness/activeness. A similar relation can be seen between vocalic non-prominence (<u>'s softness and <ʌ>'s lack of sharpness), the general quality of unobtrusiveness, and darkness/passiveness.

Collapses. As on the brilliance scales, the <ʌ/i> collapse and its central placement on the two energy scales can be ascribed to shared absence of both triggering features; this pair have neither the 'vital' amplitude of <a> nor the 'non-vital' rounding of <u>. In terms of perceived intensity, they are neutral, being both non-loud and non-soft; thus, they come to occupy the middle range. The collapse of <a/i> on passivity, as on darkness, can be attributed either to shared non-backness (in vitality-trigger terms, non-roundedness) or to the shared prominence of these vowels, or both. The third collapse, that of <ʌ/u> on passivity, can be explained similarly, as due to their shared non-prominence. Prominence contrast also answers the question of why, although <u> collapses with <ʌ>, it remains differentiated from <i>, <ʌ>'s clustering 'partner in crime:' the stimuli <i> and <u> belong, respectively, to opposed prominent and non-prominent pairs, and this ensures their separation. In addition, we speculate that some aspect of non-salience (in this case, already-noted particular difficulties in responding to the passivity scale) may have

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

emphasized the underlying phonemic relationship between <u> and <ʌ>, realized as their common non-prominence.

Data provided only by non-salient scales. One interpretation of the results examined so far in Chapters 7 through 9 is that the non-salient scales, especially the 'blunter' ones (those yielding larger numbers of collapses, more extensive chaining, ranking asymmetries, lesser robustness of response), were simply less efficient measures of phonetic symbolism. However, it is also possible to see the non-salient scales as providing different data from those of the salient scales. Because all three non-salient terms may have been more confusing to apply (especially those for the energy and size dimensions),[70] rather than focusing on subjects' attention on easy-to-grasp contrasts, they may have directed it towards those triggering features shared by a stimulus pair, resulting in approximately 50-50 scores for both vowels. Thus these scales would have tended to emphasize what the stimuli were felt to have in common phonosymbolically.[71] For example, we can see the Japanese passivity results as indicating, rather than simply confusion, a 'long view' i.e., a relaxing of distinctions between the prominent, hence non-passive, <a> and <i>, and a similar relaxing for the non-prominent, hence passive, <u> and <ʌ>. It is also easy to see how non-salient scales, with their tendency to be responsive to larger groupings, might be more sensitive than the salient scales to the generalized prominence pairings.

---

[70] See below, Section 9.2.3 (*Size compared with brilliance and energy*), where it is argued that from configuration and response range data, brilliance appears to be an easy dimension for judgments, regardless of salience, while energy, although an easy dimension, is rendered difficult by non-salience. In contrast, size is a difficult dimension, regardless of salience. These tendencies have been found in all three data sets so far examined, but are especially pronounced in Japanese.

[71] For discussion of this point, see Chapter 7, Section 7.2.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

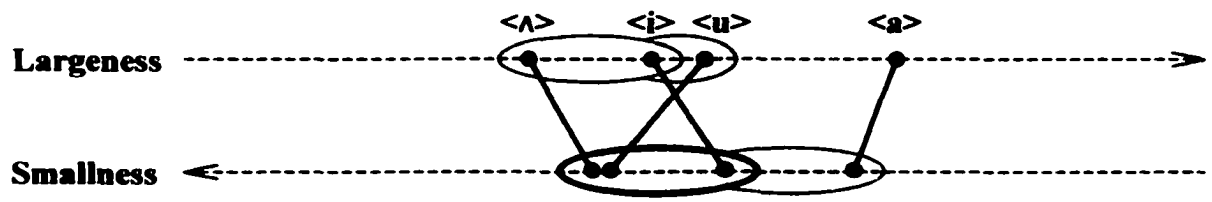
Indeterminacies on the passivity scale. The configuration here is of two chains, formed of two contiguous collapses. First is <ʌ>-<i>-<a> toward the non-passive end of the scale, and second <u>-<ʌ>-<i> towards the passive end; the middle stimulus in each chain is placed indeterminately between the outer members.The indeterminacy of <i> can be explained in terms of common features: <i> shares non-backness/non-roundedness with <ʌ> and <a>. It shares prominence with <a> (which would draw it toward the non-passive end of the scale) while with <ʌ>, it shares non-amplitude: neutral acoustic intensity and non-openness (pulling it towards the middle of the scale). We also note that <i>, in having prominence while lacking amplitude, displays a feature conflict, and its indeterminacy can be seen as reflecting this dual pull. Finally, contrast in prominence (in this case, amplitude) maintains the differentiation between the outer members <ʌ> and <a>, so that <ʌ>-<i>-<a> form a chain and not a full, three-way collapse.

The indeterminacy of <ʌ> can be explained likewise in terms of common features: <ʌ> shares non-amplitude (non-loudness, non-openness) with both <u> and <i>. It shares non-prominence with <u> (which would draw it towards the passive end of the scale), while with <i>, it shares non-backness/non-roundedness (moving it towards the middle of the scale). Since <ʌ> has neither prominence nor amplitude, it cannot be claimed to display a feature conflict. Rather, its indeterminacy reflects a compromise between its non-prominence and its tendency to be confused with (the conflicted) <i>. Lastly, contrast in prominence (in this case sharpness) again maintains the differentiation between the outer members, <u> and <i>, so that here too <u>-<ʌ>-<i> form a chain and not a three-way collapse.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

9.2.3 Size dimension

Figure 9.5 presents the data arising from judgments on the size scales by the Japanese speakers.



Largeness

	<u>	<a>	<i>
<ʌ>	65 / 34	79 / 20	53 / 46
<u>		66 / 34	46 / 53
<a>			37 / 63

Smallness

	<u>	<a>	<i>
<ʌ>	47 / 52	30 / 69	41 / 56
<u>		31 / 68	41 / 58
<a>			57 / 42

Figure 9.5, Size dimension data arrays (Japanese speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing significantly, p>.05; dotted lines indicate trend, .10>p>.05; bolding indicates a special case with no significant contrasts among a set of *three* stimuli.

General remarks on the scales. The upper panel of Figure 9.5 shows a clear tendency to nondifferentiation on both scales. The only independently-ranked stimulus is <a>, and then, only on the salient scale. Here, we find rankings are collapsed into three steps, with <i> placed indeterminately between third-largest and second-largest steps to form an <u>-<ʌ>-<i> chain, largeness being the only salient scale in the Japanese data to display chaining. Consistent with the trend to more blurred responses on the non-salient scale, collapses on smallness are even more

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

extensive. Here, we find a two-step configuration, with the triad of small stimuli <ʌ/u/i> fully collapsed; in addition, <u>-<i>-<a> form a chain, with <i> indeterminate between the collapsed triad and non-small <a>. We also note that the rankings of <i> and <u> are different on the two scales, i.e., asymmetrical, although <i> remains consistently collapsed with both <ʌ> and (on this dimension) <u>. Though not as small as for passivity, there is a restricted spread of ranking index values for both scales, resulting in fairly narrow index ranges, with a comparatively small disparity between them (of .106). The lower panel of 9.5. indicates a somewhat greater robustness of response for the salient scale, but the contrast between the two scales is less sharp than that seen on the energy dimension. On largeness, we find definitely greater robustness of response in the scores of three pairs: <ʌ/a>, <a/i> and <ʌ/u>. On the other hand, we see similar score values in favor of smallness for <u/i>, while for <u/a> and <ʌ/i> there is a more modest difference in response, also in favor of the non-salient scale.

Size compared with brilliance and energy The endpoint placement of <a> is consistent over all three dimensions; it is independently ranked on the salient scales, and collapsed with <i> as the outer member of a chain on the non-salient scales. On the other hand, <u> shows a ranking change between size and vitality: placed at the opposite endpoint from <a> on the vitality scales, it moves to the mid-range on the size scales (where we find instead <ʌ> at the other extreme from <a>). In addition, we see a further mobility for <u>, since it takes part in an asymmetry on the size scales, as both second-largest and second-smallest stimulus. Another contrast between size and vitality is the unique collapse of <i/u> on the size scales. We also note the collapses of <ʌ/u> on the non-salient energy and size scales (those with the most blurring), indicating the influence of salience, as well as dimension, on differentiability.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9.22: Japanese and Korean response range and discrepancy values for salient and non-salient scales of three semantic dimensions

	Korean			Japanese		
Dim.	*Sal.*	*Disc.*	*N-sal.*	*Sal.*	*Disc.*	*N-sal.*
Brill.	.591	.029	.562	.530	.115	.415
En.	.541	.198	.343	.541	.343	.198
Size	.436	.044	.392	.359	.106	.253

To underline, via comparison, the particular effects of salience in the Japanese data, Table 9.22 reproduces that portion of Table 9.21 which gives the Korean and Japanese range values and salient vs. non-salient discrepancies between matched scales: on all three dimensions, these discrepancies are considerably lower for Korean than for Japanese. The Korean data indicate that scales fall into two groupings of three; similarly wide range values suggest that brightness, darkness and activity were easier to apply as phonosymbolic measures, while correspondingly narrower ranges suggest that largeness, smallness and passivity were more difficult. In contrast, the Japanese range values fall into three groupings of two scales, suggesting three levels of ease/difficulty of application. Easiest/least difficult were activity and brightness; in the mid-range were darkness and largeness; least easy/most difficult were smallness and passivity.

This patterning suggests that salience had a much stronger effect on judgments for the Japanese than the Korean group (where only energy is affected, with the non-salient scale joining the 'difficult' triad, while its salient partner remains in the 'easy' triad).[72] The Japanese data suggest a more pronounced tension between dimension and salience: this is best illustrated by comparing

[72] In surveying all three data sets (see above, Section 9.2.), we have already remarked the interlinguistic tendency of passivity to elicit a very narrow response range.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

this group's scales of largeness and darkness. Besides similarities in response range values, the two scales are similar in configuration type (one independently-ranked stimulus, the 'most-X,' vs. a chain). Thus, we find salient-scale evaluations on a difficult dimension resembling non-salient-scale evaluations on an easy dimension. This indicates that although brilliance seems generally to have been an easy phonosymbolic measure to apply,[73] for the Japanese subjects, judgments on *any* of the non-salient scales may have been especially problematic. Consequently, where a scale had 'two strikes' against it—difficulties with dimension *and* salience—we might expect the maximal amount of blurring in the Japanese data, and this is indeed what we find with passivity and smallness. Besides having fairly similar range values, these scales are configured with the same absence of independently-ranked stimuli, but the collapsing on smallness (the non-salient scale of the most difficult dimension) is even more extensive than on passivity. On smallness, nondifferentiation affects four pairs, compared to three on passivity, with the result that the <u>-<ʌ>-<i> chain on passivity condenses further into the three-way <ʌ/u/i> collapse on smallness.

Interpretation of stimulus rankings. Given the generally collapsed nature of the Japanese size configurations, we see that almost any question about one stimulus will tend to involve some reference to two, or even three, of the others. The question arises here whether such complex configurations could be the (unlikely) result of a single dominant size trigger, or the product of a largely harmonious interaction between those features of prominence, amplitude and rounding argued to be responsible for the vitality rankings. A third, more likely, possibility is that, like the AE and Korean size configurations, they are caused by interactions (harmonious or conflicting)

[73] For discussion of this issue in the AE and Korean data, see Chapters 7 and 8, Sections 7.2.3 and 8.2.3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

between a number of size-connoting features (such as F_0 and perceived duration); these may include the vitality triggers. In addition, any of the vowels may have had internal interactions between two or more of its features, producing conflicts (as was suggested above in accounting for <i>'s indeterminacy on passivity). This is an issue to be raised when looking at the mid-range placements, most especially the ranking asymmetry of <i> and <u>.

In the previous discussions of AE and Korean size data, tables were provided which gave rankings of five potential size-connoting features: articulatory spaciousness, perceived duration and intensity, intrinsic pitch and locus of energy concentration. A parallell table, but with six features[74] is provided below for Japanese, for two reasons. It may serve as a tool in helping to interpret these complex configurations, especially if feature conflicts appear to be present; moreover, its parallellism may be useful in subsequent cross-linguistic comparisons of triggers. However, the reader is warned that the Japanese size trigger table is less satisfactory than the Korean one, (itself less satisfactory than that for American English). Firstly, articulatory spaciousness (which joined horizontal and vertical magnitude as sub-features of the same feature) was illuminating for the AE and Korean <a/u> size collapses (i.e., both as large/non-small). On the other hand, horizontal magnitude appears less likely to have connoted largeness for <u> in the Japanese data, since this stimulus is placed in the mid-range (differentiated from <a>, collapsed on both scales with <i> and with <ʌ> on smallness). Thus, it would be hard to ascribe a definite spaciousness value to <u>. Secondly, while values for Japanese [ɯ] have been

---

[74] The sixth is the Japanese-specific potential size trigger of devoiceability.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

located, few acoustic data have been found for the fronted,[75] and none for the backed allophones of /ɯ/, [ɯ̟], and [ɯ^{w}], argued to be the L1 models respectively for <ʌ>and <u>. In consequence, any acoustic-feature-based rankings for these vowels are speculative, and *all* orderings of the four stimuli remain tentative. Thirdly, for [ɯ] and [i], the two F_0 studies located gave conflicting values;[76] the high vowel stimuli therefore cannot be ranked according to intrinsic pitch. Fourthly, the intensity-based rankings proposed in the previous section (which refer to jaw and lip position, i.e., amplitude and rounding) are not consistent with those given for [ɯ] and [i] in the one, single-informant study found to provide intensity values (Homma, 1973). And fifthly, the table cannot take prominence pairings into account, since they are based on groupings of disparate features (for <a> and <i>) and on allophonic relationship (for <u> and <ʌ>). Yet, given the extensive collapsing on these scales, the feature of prominence may turn out to yield better explanations than those relating to potential size triggers.

Taking into account, therefore, data gaps and other unsatisfactory apsects, Table 9.23 below offers tentative size-trigger rankings. Where possible, these are supported by values taken from the studies surveyed in Section 9.1.2, with the exception of the perceived intensity rankings; these have been ordered as argued for in Section 9.2.2, rather than following the order given by Homma's (1973) data.[77] As in the equivalent tables for AE and Korean, the size rankings obtained in this experiment (with values 1 through 4) are presented for comparison with the six

---

[75] Homma (1973) provides F_1 and F_2 values for [ɯ̟].

[76] See above, Section 9.1.2, also Homma (1973) and Whalen and Levitt (1995).

[77] This choice was made to be consistent with the previous interpretation of vitality rankings as partly triggered by perceived intensity, due to lip/jaw position. In addition, the minimal size of Homma's sample makes it unreliable in ranking vowel intensities.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

size-trigger rankings; where given, these are valued 1 through 3 (maximum). For the articulatory trigger of spaciousness, this is because <ʌ> is considered as having equivalent anteriority and height with <i>, and thus an equivalent, lowest, value, which would yield a three-value ranking. However, because of the above-mentioned difficulty in ascribing a spaciousness value for <u> (distinct from that for <i> and <ʌ>), a separate, higher value for <u>, and hence a higher value for <a>, are marked only in parentheses. For the acoustic triggers of perceived duration and devoiceability, the same value is tentatively marked for <ʌ> and <u>, because of the proposed derivation of both from /ɯ/ and the above-mentioned lack of acoustic data for its fronted and rounded/backed allophones.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9.23: Observed stimulus ranking for size dimension (order being specified as *least size* = 1 through *most size* = 4), compared with rankings

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Observed Rankings				
	1	**2**	**3**	**4**
Largeness	**<ʌ>**	**<i>**	**<u>**	**<a>**
Smallness	**<ʌ>**	**<u>**	**<i>**	**<a>**
Feature-based Rankings	**<ʌ>**	**<i>**	**<u>**	**<a>**
Articulatory spaciousness	1	1	1 (*or2*)	2 (*or3*)
Perceived duration (1-mora)	1	2	1?	3
Perceived intensity	2	2	1	3
Intrinsic pitch	?	?	?	1
Energy concentration locus	?	1	3	2
Devoiceability	3 (*or 2*)	2	3 (*or 2*)	1

The rankings of <a>. Inspection of Table 9.23 suggests that <a>'s extreme size placements, and its unique independence on the largeness scale, are due to the cumulative effect of endpoint rankings, with values of 1 or 3 (or 2 on spaciousness) for five out of the six size triggers. The vertical spaciousness of <a>'s wide jaw opening has an obvious iconic connection to large size,[78] as does its high perceived intensity. Moreover, Japanese [a] has the longest intrinsic duration at either mora-length (Strange et al. 1998), and, whatever the language, [a] has the lowest intrinsic F_0 of the point vowels (Whalen and Levitt 1995). All of [a]'s acoustic features are naturally associable with largeness: large objects and creatures tend to make louder, lower-pitched sounds

[78] For early discussions establishing this point, relating to the gestural roots of phonetic symbolism, see Paget (1927, 1930) and Sapir (1927).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

than small ones,[79] and, as we have seen, an iconic connection can be made between phonetic length and phonosymbolic size.[80]

In addition, Japanese [a] offers an extra, language-specific contrast with the high vowels, in that it rarely devoices in devoiceable environments.[81] This appears to be a phenomenon largely based on the greater length of [a] compared to the high vowels: we might put it that for the Japanese speaker, [a] has more temporal 'substance' than [i] or [ɯ], and is thus less likely to become 'invisible' in terms of voicing. It is easy to see how [a]'s resistance to devoicing might translate in terms of both size *and* vitality, as large rather than small, and bright/active rather than dark/passive. Finally, apart from its mid-range locus ranking (which could be claimed to indicate size neutrality, rather than smallness), <a> shows no potential for feature conflicts; i.e., it lacks characteristics which could serve as smallness-triggers and thus cancel out the effects of its openness, long perceived duration, loudness, low pitch and resistance to devoicing.

All of <a>'s largeness features can be derived from its amplitude, with the exception of prominence, a general feature that includes amplitude; thus, it appears that earlier observations connecting <a>'s amplitude and prominence with its endpoint rankings on the vitality scales are relevant also to its size rankings. Moreover, based on the uniqueness of <a>'s independent placement on largeness, in terms of spaciousness it can be claimed that only vertical and not

---

[79] These are more properly issues of sound symbolism, rather than of phonetic symbolism. However, the argument for <a>'s largeness is strengthened, since the size-attribution will be in addition pre- and extra-linguistic.

[80] For discussion of this issue, see Chapter 7, Section 7.2.3.

[81] For details of devoicing see above, Section 9.1.2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

horizontal magnitude (i.e., amplitude alone) is the real size trigger. Thus, there are grounds for assigning the same articulatory spaciousness value (of 1) to all the high vowel stimuli including <u>, and assigning the higher value (of 2) only to <a>.

On this dimension, the only case where <a> is not differentiated from another stimulus is in the collapse with <i> on smallness. Since <a> contrasts with <i> in amplitude, their collapse here must be ascribed either to a different feature, or one that could include amplitude in a shared, larger grouping. Just such a grouping is found in prominence, i.e., the feature equivalence of <i>'s high locus/sharpness with <a>'s amplitude. It has been argued that smallness, like the other non-salient scales, was particularly apt for registering the more general phonosymbolic groupings.

The rankings of <ʌ>. The Japanese ranking of <a> as largest/least-small is intuitively understandable and agrees with its placements on the English and Korean size scales. On the other hand, the ranking of <ʌ> as smallest/least-large (rather than <i>, as in the other two data sets) seems to require more explanation. However, <ʌ>'s placement has only numerical, and not statistical, support; it collapses here (as everywhere else) with <i>, and also collapses with <u> on smallness. We must therefore remember that these rankings for <ʌ> may be the product of chance, and may mask a partial agreement with the AE and Korean size evaluations.[82]

[82] We note that <ʌ> is ranked next to <i> (as second-least-learge/second-smallest) on the AE and Korean size scales, suggesting that just such an agreement exists.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Given the large number of collapses for <ʌ> on the size dimension, our first question concerns the triggering feature(s) likely to have distinguished this vowel from <a> on both scales. As Table 9.23 indicates, a good candidate is <ʌ>'s lack of articulatory spaciousness, vertical and horizontal, compared to <a>'s vertical spaciousness; additional triggers might be contrasts in the vowels' perceived durations and in their propensities to devoice. Disagreement between perceived intensity rankings and actual size rankings on Table 9.23 indicate that perceived intensity would not have been a dominant size trigger for <ʌ> or the other high vowel stimuli.[83] Otherwise, rather than the reverse, we would have found <u> (the softest vowel, valued as 1) at the non-large/small endpoints and <ʌ> (collapsed with <i> as neither loud nor soft, valued as 2) in the middle range, as seen on the vitality scales.

Next, we ask about <ʌ>'s differentiation from <u> on largeness, and again the most likely feature appears to be articulatory spaciousness, but in this instance, it would be the contrast between the horizontal spaciousness of <u> and its absence in <ʌ>. If this interpretation is correct, then the salient scale is registering finer distinctions than does the non-salient scale (since elsewhere only vertical spaciousness appears to mark for size). Thus, on smallness we find <ʌ> collapsed with <u>: here, the most likely triggers are shared absence of vertical spaciousness and shared non-prominence (which may have included shared devoiceability and shared perceived (short) duration[84]).

---

[83] However, this does not rule out perceived intensity as a contributing factor in the ranking of <a> and in the collapse of <ʌ/i>. For arguments that subjects may well have compared different stimulus pairs on the bases of different features, see Chapter 7, Section 7.2.3.

[84] Descriptions of epenthesis, a function likely to be performed by the shortest of the vowels, support the claim that both <ʌ> and <u> (i.e., both allophones of /ɯ/) may have had the shortest perceived durations of the four stimulus vowels: see above, Section 9.1 and Akamatsu (1997). For details of vowel duration, see Han, (1961) and Strange et al.'s (1999) one-mora rankings.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In accounting for the collapse of <ʌ/i> on both size scales, their shared absence of vertical and horizontal spaciousness is again an obvious candidate. It is possible that shared devoiceability also contributed to the collapse, and shared perceived (neutral) intensity may have reinforced it. However, as noted above, if perceived intensity had been a dominant size trigger, it would have produced a mid-range ranking for this pair, as it is argued to have done on the vitality scales.

Finally, in accounting for the three-way collapse of <ʌ/i/u> on smallness, we note that all three vowels lack vertical spaciousness. In terms of their F_0's although we have no data to differentiate between them, it is expected (Whalen and Levitt, 1995) that all three had higher fundamentals than <a>. Similarly, all three are likely to have a shorter perceived duration than <a>, can be categorized as non-loud and may share similar propensity to devoice. At this point it is worth further exploring the idea that <ʌ> has phonosymbolic connections with both of the stimuli with which it is collapsed.

Phonetic vs. phonemic alleigances of <ʌ>. Our hypothesis about the pronunciation of <ʌ> and <u> for Japanese subjects has argued that the stimulus vowels, realized as four phones, are underlyingly three phonemes, /a/, /i/ and /ɯ/. Categorical differentiation can be expected between any phoneme pair (for example, the three Japanese phoneme pairs /a~i/, /i~ɯ/ and /a~ɯ/), since each phoneme distinguishes lexical items via contrast with its pair member. However, this does not mean that for any two phoneme pairs the same degree of phonosymbolic differentiability can be expected between their members (for example between the members of /a~i/ and those of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

/a~ɯ/). So far, findings have indicated that differentiability is gradeable and context-driven; it appears to be determined by one or more of four factors—language group, stimulus pair, semantic dimension and salience.[85] In the case of <ʌ> for the Japanese subjects, it can be argued that non-salience (except on the easy dimension of brillliance) had an especially blurring effect on pair judgments with <u>.

The Japanese <ʌ/i> cluster of collapses have been claimed to indicate that for this group <ʌ> cannot be distinguished phonosymbolically from <i>, due to crucial phonetic similarities between the vowels' pronunciations. But we also find <ʌ/u> collapses on the two 'bluntest' scales, likewise suggesting inability to differentiate between the members of this pair, and thus some link between the vowels. Besides these collapses (argued as due to non-prominence, deriving from shared allophonic membership of /ɯ/) we find other data patterns in which both <ʌ> and <u> respond in the same way, by consistently opposing <a>. The pairs <u/a> and <ʌ/a> are differentiated from each other on all six scales and are the only two out of the six stimulus pairs to show total differentiation. This further supports the hypothesis that both pairs are underlyingly /ɯ~a/. It can therefore be argued that data patterns for <ʌ> suggest a double alliegance: overt-phonetic (cluster-collapsing with <i>) and underlying-phonemic (patterning with <u> in the differentiation from <a> and collapsing with <u> on darkness and passivity).

---

[85] Thus, although in AE, Korean and Japanese, /a/ and /i/ are in phonemic opposition, we see different patternings of response to the stimulus pair <a/i> in the data of each group. For both Korean and Japanese, we find collapses of this pair, though not with the same distribution. In the Japanese results, <a/i> collapses on all three non-salient scales, suggesting for these speakers this pair is especially sensitive to salience, while in the Korean results <a/i> collapses on both brilliance scales, indicating for these speakers the particular sensitivity of this pair to dimension. The AE speakers, in contrast, show differentiation between <a> and <i> on all six scales.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In sum, due to phonetic and phonological/phonotactic features, <ʌ> is likely to have been perceived by Japanese subjects as lacking in articulatory spaciousness (crucially in vertical spaciousness); as being non-loud and high-pitched; as being, in the context of non-salience, indistinguishable from the non-prominent <u> to which it is allophonically related. It is also possible that compared with <i>, <ʌ> may have had a shorter perceived duration, and a greater liability to devoice, thus more tendency to 'invisibility.'[86] It is easy to see how all of these characteristics could be translated by subjects into phonosymbolic attributions of non-largeness/smallness. Finally,<ʌ> (like <a>) appears to have no feature conflicts in terms of size, i.e., no largeness-features which could neutralize the 'minimizing' effects of its closeness and frontness, shortness, softness, high pitch and devoiceability.

The rankings of <i>. The collapse of <i> with <a> has already been accounted for above, as has its collapse with <ʌ>. Therefore, we now turn to the collapse of <i> with <u> on both size scales and the asymmetry in the rankings of these two stimuli. Once again, an obvious feature likely to have triggered the failure to differentiate between these stimuli is their shared lack of vertical spaciousness; another might be their shared propensity to devoice. Since the <u/i> collapse is, unusually, found on the salient, as well as the non-salient scale and is found only on the size dimension, the ranking asymmetry of these stimuli can be considered as a particularly intense case of non-differentiation, and the rankings most probably due to chance. Given the dimension-

---

[86] For details, see Han (1961); Isshiki (1965) also suggests [ɯ] has a greater propensity to devoicing, but without overtly claiming so. Devoiceability is the only feature that may have distinguished <ʌ> (and <u>) from <i>, since it appears that of the two Japanese high vowels, [ɯ] is the more predisposed to devoice, a characteristic probably based on this vowel's notable shortness. It adds to [ɯ]'s (and hence <ʌ>'s) features a special tendency to vocal 'invisibility,' (interpretable as smallness, nonlargeness: this also suggests an iconic relation), and opposing this stimulus to <a>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

specificity of this collapse, it is worth looking for additional features, or interactions between features, which might be responsible.

Shared feature conflicts. It can be argued that in the context of size (perhaps especially on the more difficult scale of smallness) both <i> and <u> display potential feature conflicts. In the case of <i>, its prominence (high locus, sharpness) moves it rightwards to approach the non-small <a>, while its non-spaciousness (horizontal and vertical) moves it in the opposite direction, towards the least-large/smallest <ʌ>. In the case of <u>, its horizontal spaciousness on largeness ranks it as larger than <ʌ>; on the other hand, its non-prominence (shared allophonic relation to /ɯ/, propensity to devoice, perceived shortness) makes it of similar smallness to <ʌ>. Although their internal conflicts are different, the resulting ambiguity of both stimuli with regards to size may be an additional reason for their collapse.

General conclusions on the size findings. Since all details of the placements of the fourth stimulus, <u>, have already been covered in discussions of the rankings of the other three stimuli, we now pass to a brief conclusion on the size data.

The feature of vertical spaciousness will account for most of the configuration details of these scales. It will explain the rankings of <a> on both scales, and its independence on largeness; the rankings and collapse of <ʌ/i> on both scales; the collapse of <u/i> on both scales; the collapse of <ʌ/u> on smallness; and, consequently, the <ʌ/u/i> collapse on smallness. It may be that the real triggering feature is not simply vertical spaciousness vs. lack of vertical spaciousness, but

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

amplitude vs. lack of amplitude, i.e., with the feature of perceived intensity added. But if so, for the size dimension, the three-way distinction between soft (damped/rounded) <u>, neutral-intensity <ʌ> and <i> and loud <a> has become collapsed into a simple contrast between non-loud <ʌ>, <i> and <u> and loud <a>.

The one collapse that vertical spaciousness will not account for is that of <a/i> on smallness; this has been ascribed to prominence. Since it occurs only on the non-salient scale, this suggests that prominence is a less effective trigger than vertical spaciousness. We also note that two stimuli, <a> and <ʌ>, are unconflicted: all of their features can be claimed to harmonize with attributions of size. However, the other two stimuli, <i> and <u>, can be seen as having internal conflicts between size-connoting features, and it has been argued that their mid-range placements and asymmetry in rankings reflect these conflicts.

9.2.4 Concluding remarks on the Japanese data and a phonosymbolic profile of the stimulus vowels according to Japanese subjects.

Salient vs. non-salient robustness of response. The scatterplot (Figure 9.6, below), which corresponds to Figures 7.6 in Chapter 7 and 8.5 in Chapter 8 allows us to see whether the greater robustness found in the AE and Korean salient scale responses is also found in the Japanese data.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

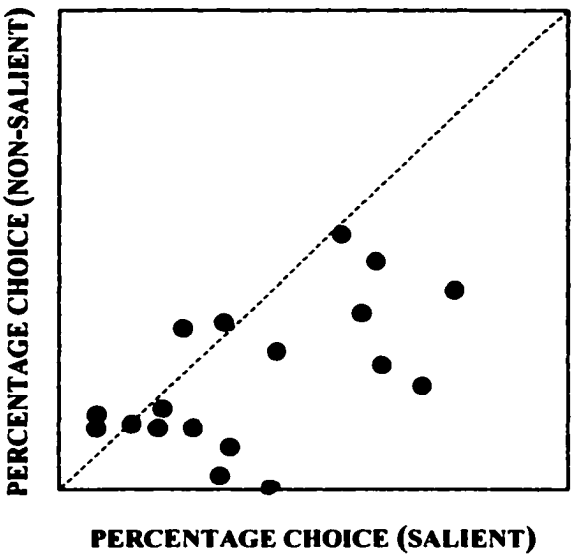


Figure 9.6, Preference strength in antonymic scales (Japanese speakers). Percentage of elective choices ('neither' responses omitted) favoring preferred stimuli for each of 6 stimulus pairs evaluated on salient vs. non-salient sub-scales of 3 dimensions. Axis values run 50 to 100%.

We note that 11/18 points are below the diagonal (indicating numerically stronger salient scale preferences for a stimulus); located on the diagonal, we find 3/18 points (indicating that preferences are of equal strength on salient and non-salient scales), and above the diagonal there are 4/18 points (indicating stronger non-salient scale preferences). Not only do we find a majority of salient scale preferences, but 7/11 salient points are located well below the diagonal while 4/4 non-salient points are close above it. Thus we see smaller majorities for the preferred stimuli on the 'winning' non-salient scales, and larger majorities for those on the 'winning' salient scales. It can therefore be claimed that the tendency found in the AE and Korean data for a more robust response pattern on the salient scales is clearly repeated here, and this agreement between the three data sets so far examined suggests a general trend.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Triggering features. Once again, as with the AE and Korean data, the rankings are the same for brilliance as for energy; the same two questions raised in the previous cases are relevant here. Firstly, were brilliance and energy synonymous for the Japanese speakers? Secondly, were the same triggering features operating in evaluations on both dimensions? A lightly qualified "yes" would be the answer to the first question: salience has a generally stronger 'blurring' effect on Japanese responses than it does on those of the other two groups, and it has a particularly strong effect on the energy dimension. But although on the non-salient scales we see definite range-shrinkages and greater tendency for stimuli to cluster together in the middle of the scale, there is still consistency in the data: from one dimension to the other, we do not see very different evaluations for any of the stimuli. Thus, for these data it is reasonable to continue considering brilliance and energy as subdimensions of a metadimension, and the use of the term 'vitality' to cover both subdimensions can again be justified.

The same observation extends to the triggering features (vitality triggers), found to be operating on both dimensions. Those proposed were, firstly, *amplitude* and *rounding*, both referring to gestalten of linked phonetic characteristics; because of their distributions, they yield together a gradeable, three-level system for the four stimulus vowels. Secondly, an additional, more general feature of *prominence* has been proposed, which includes amplitude and high locus/sharpness within it, and which yields a two-level system of grouping the stimuli (into two pairs).

Finally, on the size dimension, *vertical articulatory spaciousness* has been suggested as the main trigger (although this may be no more than a collapsed version of amplitude, with two levels of perceived acoustic intensity instead of three). Prominence is also proposed as a size trigger, but

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

considered to have less effect than it does on the vitality rankings. However, because of lack of acoustic information in several areas (such as F_0 values, values for allophones of /ɯ/), the possible effects of other potential size triggers could not be fully explored, and all interpretation of the Japanese size results must remain somewhat problematic.

Phonosymbolic profiles of the stimuli.

<a> Non-consensus: 3/18 (17%). This stimulus occupies a consistent endpoint on 6/6 scales. Its differentiability is somewhat affected by salience: it collapses with <i> on all three non-salient scales, but these are the only collapses in which it participates.The phonosymbolic personality of <a> appears to be the most separate and well-defined out of the four. It can be described as vital (bright/active) and large.

<i> Non-consensus: 11/18 (61%) This stimulus is mid-ranked on 6/6 scales. Its differentiability is affected by pair partner (cluster-collapsing with <ʌ> on all scales), salience (collapsing with <a> on all non-salient scales) and dimension (collapsing and participating in a ranking asymmetry with <u> on size). Consequently it can be described as ambiguous or unstable, or as having a phonosymbolic personality with conflicting tendencies; it appears to be the most poorly-defined of the four stimulus vowels. It is best characterized as fairly vital, but with some tendencies towards neutrality, neither very large nor small but with tendencies in both directions.

<u> Non-consensus: 4/18 (22%) This stimulus is ranked at consistent extremes on the vitality scales (independently ranked on three out of four), but mid-ranked, with an asymmetry, on the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

size scales. Its capacity to be differentiated varies with difficulty (collapsing with <ʌ> on passivity and smallness: these are proposed as the 'difficult' subset of the non-salient scales) and with dimension (collapsing with <i> on size). Next to <a>, it appears that <u> has the most well-differentiated, phonosymbolic personality. It can be described as non-vital (dark/passive) and rather more small than large, but with some size ambiguity.

<ʌ> Non-consensus: 8/18 (44%). For this stimulus, we see a contrast between a mid-range position on the vitality scales (tending towards the non-vital ends) and an extreme ranking for size, although since it is collapsed with <i>, the ranking may be accidental. Its differentiability varies with pair partner (collapsing with <i> on all scales) and with salience/difficulty (collapsing with <u> on passivity and smallness). In terms of having a defined phonosymbolic personality, it appears less ambiguous than <i>, but not as well-defined as <u> or <a>. It can be characterized as fairly non-vital and small.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 10

ANALYSIS AND INTEPRETATION OF THE SPANISH DATA

10.1. The phonetic/phonemic background of the subjects

10.1.1 Articulatory details

Unlike the larger English and Korean monophthongal systems, but like the Japanese (to which it is sometimes compared) Spanish has only five vowels; as with Japanese, there is no Spanish phonemic equivalent to AE [ʌ].The Spanish point vowels, [a], [i] and [u], more or less correspond in tongue height, anteriority and lip gesture to the stimulus vowels <a>, <i> and <u> (targets, AE [a], [iy] and [uw]). Yet anyone comparing the two languages, e.g., learning Spanish or teaching English to Spanish-speakers, will hear the qualities of Spanish and AE vowels as quite different, whatever the correspondances between the two systems. In articulatory terms, these differences in quality are ascribable chiefly to tongue positions which are more extreme and more stable, without the English-speaker's tendency to offglide (Navarro Tomás 1967; see also Sorzano Jorrín 1928).[1]

[1] In Navarro Tomás' *Manual de pronunciación española* (1967), the standard work on the phonetics and phonology of Peninsular Spanish, a noticeable difference is described between the "clear and definite" character of Spanish front and back vowels and those of English and Portuguese. As an example of this, the English back vowels are cited as not having the true velar timbre which is given by the marked tongue retraction and raising found in Spanish pronunciation (Navarro Tomás 1967: 36). The stability of Spanish

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The point vowels. Navarro Tomás characterizes Spanish [i] as articulated with "the tip of the tongue lodged against the lower front teeth; the center[2] is raised against the hard palate, touching it fully on both sides and leaving a relatively narrow opening in the middle; this contact generally reaches in front as far as the canines...[There is] a widened opening of the lips with the corners slightly pulled back.; muscular tension [is] medium.. [This vowel is] less closed or tense than the German, French or English [high front vowel] but quite close to them, above all in strong pronunciation" (Navarro Tomás1967: 46-47). Serralta and Tusón (1967) agree in describing Spanish [i] as less tense than its French counterpart, as does Colton (1909),[3] who, however, still classifies its long allophone[4] as a tense vowel. Sorzano Jorrín (1928: 89), describes Spanish [i] as less tense than English [iy] but not as lax as [ɪ].

Compared to English [uw], Spanish [u] has relatively strong lip rounding/protrusion, tongue raising, and lacks the offglide.[5] According to Navarro Tomás (1967: 38-39)

---

vowels is also emphasized; "whether high or low," like the French (high) vowels, they are "normally invariable from the beginning to the end of each sound." For this reason, an English-speaker's tendency to diphthongize stressed vowels when applied to Spanish gives a "strange and shocking" quality to sounds which should have a "fixed and homogenous timber... required by the phonetic character of the language" (Navarro Tomás 1966: 43-44). An earlier writer, Sorzano Jorrín, author of textbooks on the English language for Spanish speakers, when describing the pronunciation of Cuban Spanish vowels also stresses the homogeneity of Spanish vowels compared to those of English (Sorzano Jorrín 1928: 89-93).

[2] Navarro Tomás calls this part of the tongue "dorso," (literally translated as 'back'). However, from the diagram illustrating [i]'s tongue position, it is the central part of the tongue which is raised.

[3] I am indebted to Professor William Stewart for introducing me to this author and to his extremely detailed articulatory descriptions.

[4] For a discussion of allophonic variations of the point vowels, see below.

[5] Navarro Tomás (1967: 60-61) remarks that British and AE speakers tend not to round or protrude the lips enough when articulating [u] and that their attempts are likely to be too open and lacking in stability, i.e., there is a reduction in tongue height and a tendency to a [w] offglide. Sorzano Jorrín (1928: 93) also notes the absence of offglide in Spanish [u], compared to the equivalent vowel in English.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Spanish [u] has tighter lip rounding than for [o], the reverse of the Japanese case. The lips are described as being rounded and protruded, the tongue well retracted and raised toward the velum. The tip of the tongue is at the level of the lower gums, either separated from them or lightly brushing them, positioned as if suspended in the hollow of the lower jaw. As with [i], there is a "medium" degree of muscular tension. Colton (1909: 35), contrasting this vowel with French [u] rather than AE [uw], characterizes both its high tongue articulation and lip rounding gesture as "lazy," adding that part of the rounding appears to be effected by the tongue, (i.e., presumably by some form of compensatory extra retraction). Sorzano Jorrín (1928: 93-4) describes Cuban Spanish [u] as having an equal degree of closeness as [i], which suggests a slightly lower tongue position than for AE [uw], but not as low as for [ʊ].

Navarro Tomás (1967: 35) implies that the archetypal Spanish [a] unconditioned by phonetic environment is a central vowel: "The tongue position most similar to that which the tongue assumes when one breathes silently with the mouth open is that corresponding to the vowel [a]...The [a]...normally pronounced in Spanish" is articulated with "slight raising of the back part [of the tongue] toward the middle part of the mouth; the tip of the tongue is a little below the edge of the lower teeth, touching their inner surface around the gum level; the point of articulation determined by the slight raising of the back of the tongue corresponds approximately to the boundary between the hard palate and the velum, *of equal distance between the points corresponding to [i] and [u]*" Navarro Tomás (1967: 54-55, emphasis added).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sorazano Jorrín (1928: 91) describes [a] as being articulated with the tongue flat in the mouth: "Because of this, it is called a central vowel (vocal media)." He also contrasts the lip positions for Spanish and English [a]'s: "The opening of the lips for Spanish [a] is more in the horizontal direction; the opening in English [a] is more in the vertical direction... resembling an [o] which is exaggeratedly open" Sorazano Jorrín (1928: 91). Colton (1909: 23-30), who discusses Spanish [a] at great length, devoting more description to it than to any other vowel, rules out the idea of a central, 'default' [a]. Instead, he asserts that there are two [a]'s, back and front /a/-allophones, which can be further subclassified into longer and shorter variants. The opening sentence of his first chapter states simply, "Castillian possesses two [a] vowels: [a˂] and [ɐ]."[6]

Since Spanish [i] and [u] can be contrasted with their English counterparts in terms of their intense and stable tongue gestures, i.e., lack of offglides, we would predict that for Spanish-English bilinguals (like the subjects in this experiment), the English influence on their L1 high vowels would be to introduce offglides—and/or to increase duration.[7] Lip rounding for [u] appears more intense than for AE [uw], and possibly the same is true for lip spreading on Spanish [i], but as none of the authors consulted compare it with that of AE [iy], we cannot generalize about lip gesture from their observations[8] Finally, the

---

[6] The phonetic notations [a˂] and [ɐ] will be used respectively for the front and back variants of /a/.

[7] In fact, this type of hybridization of all four non-low Spanish vowels, (corresponding as they do to English diphthongs) is apparently widespread among Spanish-English bilinguals (Professor William Stewart, personal communication). A typical example, personally witnessed, is of the utterance-final glide on the mid vowel in the following: [me yamo hozey] (*me llamo José*).

[8] However, Navarro Tomás (1967: 38-39) observes that compared with rounded French vowels, the lips are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

central position of the Spanish low vowel [a] may result in a timbre closer to its American English counterpart than is found with the high vowels.

Factors influencing pronunciation of the stimuli. Description of the basic phones will not, however, answer the question of which of the allophones of Spanish [a], [i] and [u] were the most likely candidates chosen as L1 references in pronouncing the stimulus vowels. While the canonical Spanish [a] may be central, Navarro Tomás and others—most notably Colton (1909)—describe considerable horizontal variability associated with this vowel (as also appears to be found with AE [a] and Japanese [a]).[9] Moreover, in the case of the high vowels, definite variation is found on the vertical dimension (Colton 1909, Navarro Tomás, 1967, Schnitzer, 1997).

One of the key phonetic/phonological factors determining production of allophones is the type of syllable in which a vowel occurs. Thus, a second issue is subjects' interpretations of the stimulus spellings, most especially in the phonetic realizations of the final *h*'s of *hah* and *huh*. Moreover, since /ʌ/ does not exist in Spanish, the presence of <ʌ> is another complication likely to have affected allophone choices for either <a> or <u>. In determining the most likely pronunciation of the two non-front stimuli, it is necessary to decide which allophone of [a] or [u] would have been 'occupied' as the best choice for <ʌ>, and thus probably unavailable for modelling another stimulus. Finally, a third

---

more relaxed for Spanish speakers (see also Serralta and Tusón, 1967). Similarly, Navarro Tomás describes the lips as less spread for the Spanish front vowels than their French equivalents, recalling the description of Japanese vowels, where lip gesture is described as minimal (Vance 1987).

[9] This observation is made by Schnitzer (1997: 21, 66) comparing English and Spanish [a]'s. See also

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

potential influence on pronunciation is dialect variation, since the hundred subjects were drawn from eighteen Spanish-speaking countries.

Phonological environments conditioning point vowel allophones. In his *Manual*, Navarro Tomás (1967: 45-64) describes allophonic variation of the non-low vowels as conditioned by three factors and by interactions between these factors.[10] These are *syllable type* (free or checked), *degree of accent* (strong stress, weaker stress and especially weak stress) and *speech style* (slow and careful, or fast and relaxed). For [a], *the anteriority of a following consonant, vowel or semivowel* is a fourth conditioning factor, and this seems to be more important than syllable type.[11]

Three allophones are described for the non-low vowel phonemes: a more tense upper, a more lax lower, and a still more lax centralized version. For [a] there are four variants, comprising an anterior, a posterior, and a central phone (tenseness unspecified for these three), and a fourth variant which is lax and somewhat raised, its anteriority depending on the segmental environment. If occurring in a free syllable, then any of the four nonlow vowel phonemes will be realized as the higher, more tense allophone. In a checked syllable which is also stressed, the slightly lower and laxer version is produced, while if

---

Chapters 7 and 9, Sections 7.1.1 and 9.1.1.

[10] As in the case of especially weak stress position (*posición especialmente débil*), where Navarro Tomás (1967: 36) indicates an interaction between speech style and stress with each enhancing the effect of the other: "In especially weak position, between a primary and a secondary accent, and... above all... in rapid, informal speech" a shortened, lax variant of the vowel is found.

[11] Thus, e.g., central [a] is found both in the checked syllables of *reparto* and *practico* and the free syllables of *caro* and *pedazo* (Navarro Tomás, 1967: 55).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the vowel is found in "in especially weak position" (i.e., in a syllable which is neither open nor stressed nor carefully enunciated), we find the most lax and centralized allophone. In the case of the low vowel [a], as mentioned above, variation is conditioned by the anteriority of a following segment to which it assimilates progressively, or by being in especially weak position. Although [a]'s anteriority will depend upon its specific segmental environment, weak positioning will consistently raise and "weaken" it (i.e presumably pull it toward a middle articulation), producing one close in quality to the nucleus found in the (pre-World War II) Standard Southern English pronunciation of "hut" (Navarro Tomás 1966: 54-57).[12]

Other authors agree with Navarro Tomás. Describing "a somewhat idealized pattern which is rather typical of Puerto Rican speakers," (i.e. of a Carribbean, rather than Peninsular, variety of Spanish) Schnitzer (1997: 66-7), gives a much simplified version of Navarro Tomás' observations, while concurring with his general outline.[13] Colton (1909), writing on the Spanish of Castile, distinguishes two basic variants of the high vowels [i] and [u], and three of [a].[14] Like Navarro Tomás, he notes cases where different factors—

---

[12] I am obliged to Professor William Stewart for modelling this vowel for me (personal communication). He describes it as lower and farther back than AE [ʌ], articulated almost identically with Portuguese unstressed [a]; to my ears, it is quite similar to, but upwards and forwards of, the British [a] of *father*. It should be added that articulation of English [ʌ], like the Spanish weak [a], appears to vary quite considerably: it is consistently non-high and non-front, but may be central or back, mid-low or mid (see Chapter 7, Section 7.1.1; see also Ladefoged 1982 and Mackay 1987).

[13] Schnitzer (1997) notes that all four of the non-low Spanish vowels, have a higher and lower allophone, the higher one being used in open syllables and the lower one in closed syllables, while [a] can vary in anteriority.

[14] The five conditioning factors observed by Colton (1909: 24) are a) syllable type; b) the presence of an [o] in the following syllable, causing vowel mutation (*métaphonie*), as e.g., when the [o] of [paso] may have a backing influence on the preceding [a]; c) the influence of a following consonant; d) stress; e) position in the word or utterance. Speech style is not included.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

such as stress and position, or stress and syllable type—interact (Colton 1909: 30). Of the variants of [a] he remarks: "The distance between one variant and the next is not very large in Castillian, but in every case there is a horizontal difference, a vertical difference and a difference in quantity" (Colton 1909: 27).

Allophones of /i/ and /u/. A number of issues (influence of stress; paralinguistic effects; pronunciation of initial orthographic *h*; realization of <ʌ> relative to <u>) remain to be settled before we can fully conclude about pronunciations of the high vowel stimuli. However, from Navarro Tomás' (1967) descriptions of [i] and [u], it appears that the free or checked character of the syllable containing the vowel is the crucial determining factor for height of the allophones produced; stress and speech style are less important triggers of variance. This impression is corroborated by Schnitzer's (1997: 76) observation "In general, the higher allophones of the Spanish vowel system will be identified with the strong[15] English vowels." Likewise, Colton (1909: 36, 41) observes that the higher allophones of [u] and [i] are produced in open syllables, and Serralta and Tusón (1967: 94) its corollary, that the lower variants are found in closed syllables. Given the spellings of the high vowel stimuli as *hee* and *hoo* (and leaving aside the issue of stress for the moment), the vowels would always appear in the same phonetic environment, i.e., in open or free syllables, preceded by [h] and followed by either word- or phrase-boundary. Thus, it seems probable that for <i> and <u>, Spanish-speaking subjects produced either the high allophones of [i] and [u] from their own L1, or interlanguage Spanish-English

---

[15] These are the tense, diphthongized [iy], [ey], [uw] and [ow], i.e. those occurring in (English) open syllables—such as <i> and <u> (Schnitzer 1997: 68-69).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

versions. [16] These may have been longer than Spanish vowels, and even have included offglides, but are still expected to have been modelled on the higher Spanish variants.

Allophones of /a/. Of the four /a/ allophones listed by Navarro Tomás, one is the very lax and somewhat raised phone with varying anteriority which occurs in especially weak position (Navarro Tomás 1967: 57). Second is the (low, tense) central allophone (Navarro Tomás 1967: 54-55)[17] also described by Sorzano Jorrín (1928: 91).Third and fourth are the front and back allophones of [a], [a˂] and [ɐ] (Navarro Tomás 1967: 55-57), described also by Colton (1909: 23-30). Since Colton's descriptions of them are the most detailed, they are quoted below.

---

[16] I am obliged for this suggestion to Professor William Stewart (personal communication), who observes that Spanish speakers in an English-speaking environment such as NYC's may show evidence of back-and-forth phonetic interference from both languages. There are acoustic data (Peterson and Lehiste 1960: 73), indicating that even for native speakers, AE vowels divide into a long group and a short group, those with glides being among their group of "intrinsically long syllabic nuclei," while the short group consists of those vowels lacking glides. It can be argued that a Spanish speaker attempting to reproduce an AE high or mid-high vowel might either insert a glide or simply lengthen his or her steady-state L1 equivalent, in consequence changing the durational relations between the Spanish non-low vowels and [a], which are generally described as of similar length (NT, Schnitzer 1997). Thus, [i] , [u], [e] and [o] would become long, while [a], lacking a glide, would remain short.

It is true that [a] can also be offglided in AE (with [ə]), but (except in the case of a following nasal—not part of the experimental context) the offglide is less noticeable, perhaps because the [w] and [y] glides can be, and frequently are, pronounced fortis, i.e., emphatically. It is more difficult (unless one is a Southerner) to make a fortis [ə] offglide. Professor Stewart also remarks that the phonetic structure of Spanish is not inimical to offgliding vowels with [y] and [w] as it would be to offgliding them with [ə], and thus the [ə] offglide does not occur in Spanish, whereas the [w] and [y] offglides can occasionally be heard in Spanish that is not influenced by English.

[17] This allophone is considered by Navarro Tomás to be "the [a] normally pronounced in Spanish" i.e., presumably the variant closest to the phoneme /a/. See also Colton (1909: 26) who describes a similar [a] variant in intertonic and transitional position.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The first variant is the "more open" palatal[18] [a˂], "which is tense, but one will see that it is rather closer to the form attributed to the lax vowel. The tongue is flattened out in the mouth, especially in the middle and at the back. At the back, a channel is formed, and the opening between the tongue and the uvula is much greater for [a˂] than for [ɐ]. Likewise, the mouth is more open for [a˂]. The resonance chamber is spacious at both front and back. The form of the tongue is rather concave, since it is flattened against the floor of the mouth. Even so, it is barely withdrawn from the incisors.

"The tongue muscles are tense, but this tension appears even indirect or sympathetic. For the characteristic feature of this vowel is the tension in the throat and in the vocal cords. no doubt accompanied by a greater than ordinary tension in the thoracic cage. It appears that the vocal cords take up the entire width and speed of their vibrations from the beginning of the vowel, and keep these qualities right to the end. The tongue does not change position during the excecution of the vowel (whose)... timbre is rather palatal. The lip opening, while not being very large normally, is sensitive and will become exaggerated under (the effect of) emphasis, as will the other characteristic features of this vowel" Colton (1909: 23).

"The second [a]—i.e., [ɐ]—is more close. The tongue is slightly pulled away from the teeth (backwards into) the throat and is raised toward the uvula at the back. The surface of the tongue is convex, especially at the back. The tip is slightly raised, bunched and

[18] This term is challenged by Professor Edward Bendix (personal communication): "How can a low vowel, [a] have a palatal variant? What seems to be meant is a pronunctiation slighly further forward and raised than in [standard] Castillian, in the direction of, but not as far as, [æ]."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

floating, so to speak. The muscles on the inside of the mouth and throat are relaxed. The tongue muscles are just tensed enough to make the contraction. It seems to me that the tongue retracted in this way presents a softer surface to the vibrations than does the tense tongue of the emphatic [a˂]. The timber of the sound is low-pitched (*grave*) and tends toward the neutral...The acoustic effect of this vowel is very soft" Colton (1909: 24)

Thus, Spanish [a] may be realized as [a˂], a very tense, open, somewhat frontal sound, characterizeable as palatal; therefore, it might be expected to share some phonetic, hence phonosymbolic, characteristics with [i]. In contrast, [ɐ] has a distinctly more back sound[19] which is more lax and less open, i.e., with a somewhat higher tongue postition. The impressionistic term 'soft' (*doux*) in Colton's description may indicate low acoustic amplitude/intensity, or concentration of acoustic energy in the lower frequencies (as for [u]). Or it could indicate diffusion of acoustic energy (as for [ʌ]), since [ɐ] is also described as having a timbre tending toward the neutral. Consequently, [ɐ] might be expected to share certain phonetic and phonosymbolic qualities with Spanish [u], or with AE [ʌ].

Possible effects of stress. All of the evaluation questions on the questionnaire asked for pronunciation of stimuli in the following form: "Make the sounds *hah-hoo*, *hoo-hah*, *hah-hoo*, *hoo-hah*." In any or all of the questions, could the stressing of one of the stimulus

---

[19] Colton however, does not label [ɐ] 'velar' as does Navarro Tomás with the back variant of [a]. Why? Professor William Stewart (personal communication) suggests this may have been because Colton observed the articulation to be even farther back, i.e. uvular.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

syllables more than the other have had some allophonic effect? It is certainly possible that some variation in the production of individual stimulus tokens may have been caused by the way that the subjects were asked to juxtapose the two stimulus vowels in each question. If uttered in quick enough succession to mimic a bisyllabic word, the first syllable of the pair could have received main stress and the second, secondary stress. Weak stress was not a possibility (Navarro Tomás, 1967: 36). Since these are bisyllables, normal Spanish stress-assignment on the penultimate syllable, and (for those more experienced with English) normal English stress-assignment on the first syllable would have coincided.

But if this type of accenting had occurred and had some effect on vowel qualities, it should not have been fixed for any stimulus token even over a single question, let alone over all of the twentyseven in which that stimulus was presented. Any variations due to stress would have been evened out by the alternations in the order of the paired syllables, since each question required two tokens of each order. Thus, if a subject produced stress/position-related variants of the stimulus vowels, both vowels would have been pronounced in primary- *and* secondary-stressed forms.

Possible paralinguistic influences. Moreover, an argument can be made for the paired stimuli as having been as treated two monosyllables (i.e., both as having primary stress) rather than as bisyllabic words. It seems most likely that the nature and the strangeness of the experimental task would have elicited careful, rather than rapid, speech. Two vowels were being compared and evaluated according to a number of puzzling criteria; therefore,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

it can be expected that both stimuli received full focus, rather than one's being subordinated to the other. However, this point remains speculative; speech style was not explicitly controlled for, an oversight in the experimental design.[20]

The effects of word-list and stimulus spellings. Because the Japanese and Korean writing systems are quite different from the English, L1 orthographic interference was not considered an issue for the Asian language speakers. In addition, their capacity to interpret word-list and stimulus spellings satisfactorily may have been due to an already considerable literacy in the second language.[21] On the other hand, the Spanish speakers may well have suffered from confusions between L1 and L2 orthographies: the same writing system is used for both English and Spanish, but in somewhat different ways. Moreover, the Spanish speakers were generally less experienced ESL students, i.e., less likely to be literate in the second language.[22] Finally, although no data were obtained to bear out this impression, it appeared to the experimenter that the Korean and Japanese subjects were overall more affluent and middle-class than were the Spanish subjects. This suggests that the Spanish speakers may have been less sophisticated than the Asian

---

[20] A slow pronunciation with main stress might have been ensured for each syllable by changing the instructions as follows: a) Including directions on the questionnaire for subjects to *slow down* while pronouncing the stimuli b) Visually pairing the syllables in another way than with the wordlike connection of hyphenating them c) Spelling the stimuli with capitals (i.e., suggesting greater emphasis) rather than with lower-case characters.

[21] We note that 84% of the Koreans and 89% of the Japanese reported five or more years of experience studying/speaking English (see Appendix 2).

[22] 38% of the Spanish speakers reported 1 year or under of experience studying/speaking English; a further 13% had 1 to 2 years of English experience. Only 35% had 5 or more years of English experience.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

language speakers, less confident at interpreting unusual spellings, and thus more liable to misinterpret them.

The six potential problems for Spanish speakers in interpreting stimulus and word-list spellings are listed below, with vowels presented in order of ascending difficulty: 1) the initial *h* of the stimulus spellings; 2) the vowel of *hee*; 3) the vowel of *hoo*; 4) the vowel of *hah*; 5) the vowel of *huh*; 6) the final *h* of the stimulus spellings.

Initial h. Orthographic *h* in Spanish always represents [ø]; it occurs word-initially, as in *harina* [arina] 'flour' or syllable-initially, as in *inherente* [in-er-en-te] 'inherent.' In contrast, English orthographic initial h has two phonetic realizations: normally it represents [h], as in 'hurry,' but occasionally, [ø] as in 'hour,' or 'honest.' Which language's set of spelling-to-pronouncing rules were most likely to have been operating here (and if English, which of the two rules)? There are three arguments in favor of subjects' representing initial *h* as [h]. Firstly, the setting of the experiment was in English: all but one[23] of the subjects were students of English in an ESL classroom (and presumed to be already reasonably familiar with the normal correspondance of English orthographic initial *h* to phonetic [h]). Moreover, the spoken instructions, word lists and questionnaire were all in English. In addition, the stimulus spellings *hee, hoo, hah* and *huh* followed English, and not Spanish, orthographic rules—which would have produced *hi, hu* and *ha* (or *xi, xu* and *xa*) for the point vowel stimuli, and, presumably, nothing for *huh*. Secondly,

[23] The hundredth subject was a personal friend with an extensive knowledge of written English, and with fluent, fairly accurate spoken English.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

subjects had four separate opportunities to practice with [hV], since initial *h* was consistent over all four stimulus spellings. And thirdly, pronunciations with [øV] would have been noticeable and discouraged by the experimenter, those making errors being pressured to conform. On the other hand, [hV] pronunciations would have been approved.[24] Similarly, had subjects substituted their own L1 velar fricative [x] for AE [h], the resulting [xV] pronunciations would have been noticeable and discouraged.

The vowel of hee. During practice with the word lists, even the most unsophisticated subjects appeared to understand that the spelling with the digraph *ee* represented [i] (or possibly [iy]). Any Spanish-to-English orthographic confusion that might have resulted in [he] pronunciations would have been easily noticed and discouraged, but these were not evident. Instead, even though the word list included a number of different spellings other than *ee* for this stimulus vowel,[25] subjects appeared consistently quicker and more confident in obtaining a uniform pronunciation (i.e., of [hi] or [hiy]) than they had been for *hah*, the vowel practiced in the previous (first) word list.

The vowel of hoo. Very similar observations can be made about subjects' confidence and speed in pronouncing this stimulus vowel when practicing with the word lists. Once again, the spelling with the digraph *oo* seems to have been clearly understood as representing [u] (or [uw]). Spanish-to-English orthographic confusion which might have

[24] For a description of methods used to obtain uniform stimulus pronunciation without modelling the vowels, see Chapter 7, Section 7. and Chapter 9, Section 9.1.1.

[25] List items included *Lisa, Pete, cleaner* and *people*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

resulted in [ho] pronunciations were not evident, and as with the *hee* list, the assortment of spellings other than with *oo*[26] on the word list for this stimulus vowel did not appear to have confused subjects. A uniform pronunciation of [hu] or [huw] was quickly obtained.

Conclusions about the pronunciations of <i> and <u>. At this point, we can establish with reasonable certainty how the high vowel stimuli were pronounced by the Spanish speakers. Initial [h] was used (as is claimed for all four stimuli) and the higher allophones of [i] and [u]. The only remaining question is about the possible influence of AE vowels i.e., whether length and/or diphthongization were present: in other words, was the most likely pronunciation of <i> [hi], [hi:] or [hiy] and of <u>, [hu], [hu:] or [huw]? Evidence for the longer or diphthongized pronunciations' being used would be a collapse of <u/i> on the size dimension, together with evaluations as large/non-small.[27]

The vowel of hah. Out of the five Spanish vowels, the two possible L1 candidates for *hah* are the only vowels which are both non-front and non-high, [o] and [a].The word-list for *hah* presented particular potential for confusion for Spanish speakers, because of the AE orthographic representation of [a] by *o* and the inclusion of English-Spanish cognates. The list is therefore given in full below, with primary-accented syllables capitalized, as on the original document. Asterisks mark those words which correspond in cognate form (though not necessarily in meaning or pronunciation) to Spanish words.

1.) POP* 2.) LOTTery* 3.) BODy 4.) ROCKy* 5.) SOLid* 6.) MODelling*

[26] List items included *shoes*, *Lewis*, *soup* and *truly*.

[27] For a discussion of the relation established between phonosymbolic largeness and vowel length, see the discussion of Sapir (1929) and Newman (1933) in Chapter 2; also Chapter 7, Section 7.2.3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

7.) forGOTTen 8.) STOP 9.) WATCHing 10.) HOLly 11.) ROBert* 12.) PROFit

There are three arguments in favor of an [o] pronunciation for *hah*. Firstly, 11/12 list items are spelled with an *o*, and orthographic *o* in Spanish invariably indicates [o]. Secondly, 6/12 items (*pop, lottery, rocky, solid, modelling, Robert*) are cognate with the Spanish *papi, lotería, roca, sólido, modelo* and *Roberto*;[28] thus, subjects' sense that [o] was expected—at least in 5/6[29] of these words—could have been further reinforced. And thirdly, this was the first of the word-lists to be tackled, and the unfamiliarity of the task appeared to maximize subjects' confusion.[30] If the [o] pronunciation for *hah* was used, then a series of collapses for <u/a> might be predicted, due to the effects of shared rounding.

On the other hand, several arguments can be made in favor of a pronunciation with [a]. Most important is that the spelling of the target syllable, with *a*, which in Spanish invariably indicates [a]. In addition, one cognate and three of the non-cognate list items (*Pop, body, stop* and *watching*) are common American words normally pronounced with [a]. In the possible tug-of-war over [o] versus [a], it is questionnable how the more unusual words *profit* and *Holly* were pronounced, but if *forgotten* behaved like the more

---

[28] I am indebted to Dr. Ricardo Otheguy for drawing my attention (in a personal communication) to possible problems in the word-list area, especially potential [a]-[o] confusion due to cognate forms.

[29] Only in the case of *pop* (*papi*) does the cognate Spanish form require an [a].

[30] As observed above, on every occasion, obtaining correct pronunciation for *hah* took subjects a much longer time than for the second list (for *hee*).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

common *gotten*,[31] then this too is likely to have been pronounced with [a]. Finally, [o] pronounciations for *hah* would have been noticed by the experimenter and definitely discouraged, while [a] pronunciations would have been praised.

Clearly, none of the word list procedures would completely rule out individual wavering between [o] and [a], nor any subject's forgetting that the spelling *hah* was to be pronounced with [a] when the questionnaires were handed out. But although the *hah* list, as the first of the four, seemed the most lengthy and anxiety-producing to practice with, on every occasion it appeared that satisfactory collective pronunciation with [a] was eventually obtained. The question of the most likely allophone of [a] chosen for *hah* is answered in the following discussion of the most probable vowel produced for *huh*.[32]

The vowel of huh. Out of the five Spanish vowels, the three possible L1 candidates for *huh* are the non-front vowels [u], [o] and [a]. The word list for this stimulus is very likely to have confused the Spanish-speaking subjects, since, as observed above, the vowel it exemplifies has no L1 phonemic equivalent. In the English orthographic system, [ʌ], is mostly represented by *u*, but also in certain commonly-used words by *o*, and occasionally by *ou* or *oo*[33]. Once again, the word list for *huh* is given below, with primary-accented syllables capitalized and cognates asterisked.

---

[31] In my experience as a remedial ESL teacher, *forgot* and *forgotten* are most commonly spelled by Spanish speakers as *forgat* and *forgatten*.

[32] See above for arguments that the selection of one influenced the selection of the other; see also Chapter 9, Section 9.1.1 for similar discussion of Japanese solutions to the 'problem of <ʌ>'.

[33] This is a rare spelling, but found in *flood*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1. FUSS 2. SUBway 3. BUT 4. MOTHer* 5. teaCUP 6. SOME
7. acCUSTomed* 8. OTHer* 9. RUSH 10. DUSTy 11. UNder 12. COUPle

Pronunciations with [u] can be considered as a possibility, as 8/12 list items (including one cognate: *accustomed*,'acostumbrada')[34] are spelled with *u*, which invariably indicates [u] in the Spanish orthographic system. However, the [u] pronunciation is the least probable of the three. The list for ***huh*** immediately follows the ***hoo*** list and is on the same page. Thus, subjects would have just finished practicing with some variety of [hu] for the previous stimulus, which makes a repetition of the vowel unlikely. Moreover, [hu] pronunciations for ***huh*** would have been noticed by the experimenter and would have been strongly discouraged.

In contrast, pronunciations with [o] seem at first glance to be the most likely. Firstly, we note that in 3/12 list items ***huh*** is represented by *o* (invariably indicating [o] in the Spanish system). These include one cognate: *other*, 'otro,' 'otra,' while the translation of another, *accustomed*, 'acostumbrada' has *o* on the equivalent syllable. Second, and more importantly, the Spanish-influenced rendition of [o] for AE [ʌ] is very widespread.[35] It appears to be the standard pronunciation for commonly-used words where where *o*

[34] However, we note that the vowel of the corresponding syllable in the Spanish word is is not with [u] but [o] (see discussion of [o] pronunciations below).

[35] My own impressions here, based on three decades of ESL-teaching experience with Spanish speakers, and on fourteen years of living in a Dominican neighborhood (Washington Heights, NYC) are confirmed by Professors Ricardo Otheguy and William Stewart (personal communications). This does not mean that [a] substitutions for [ʌ] may not also be present, but as a teacher, I am especially listening out for [o] pronunciations , i.e., checking for this specific and widespread Spanish-speakers' error. In contrast, I will not register [a] pronunciations as requiring correction, because to my ear (conditioned by English phonology) these are acceptably close to [ʌ].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

represents [ʌ] (such as the list items *other*, *mother*, *some*, and others, e.g., *brother*, *money*, *love*, *Monday*, *come*),[36] pronounced with a laxly-rounded, lowered [o].[37] Moreover, this substitution of [o] for AE [ʌ] appears to be just as common in words spelled with *u*; this may be heard (n NYC speech) in [boŋ] 'bum,' [otauŋ] 'uptown,' and also in the word-list items *subway*, *but* and *under* (typical pronunciations being [soβe], [bo] and [onðer]). Thus, there is no reason to expect that the less frequently-used items on the *huh* list, *fuss*, *teacup*, *dusty* and *rush*, and the confusingly-spelled *couple* would not have also been pronounced with [o], particularly since it was emphasized that although spellings of the target vowel varied in words on the word-lists, all the items were to be pronounced with the same vowels in the capitalized syllables. Finally, because [o] was not specifically targeted anywhere on the word lists, using it for the pronunciation of *huh* would have kept *huh*'s vowel separated phonetically and psychologically from that of *hah*.[38]

On the other hand, precisely because the substitution of [o] for [ʌ] is such a noticeable feature of Spanish-influenced English, pronunciations of *huh* as [ho] or [hɔ]—i.e., with rounding—would have been immediately recognized by the experimenter, and would have been strongly discouraged. In sum, the feature of rounding on both [o] and [u] would have made either noticeable mispronunciations for *huh*. However, there exists a

---

[36] Professor Otheguy's observation is that where there is uncertainty about pronunciation, spelling wins the day. This is similar to Ms. Katsuno's observation about the Japanese pronunciation of loanwords, (e.g., *subway*) with [a], except for *o*-spelled words (like *London*), which are with [o] (see Chapter 9, Section 9.1.1).

[37] This is close to AE [ɔ] in quality, but in terms of Spanish should be identified as /o/.

[38] However, the use of *o* in the digraph spelling of *hoo* might then have created confusion between *hoo* and *huh* (as [ho]).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

third possibility: that this stimulus was pronounced with some variant of [a]. If so, it would not have been so evident to the experimenter that subjects were missing the target, and this suggests that [a] pronunciations may have been the most likely.

There are three pieces of evidence in favor of [a] pronunciations for *huh*. The most trivial is that one of the three cognate items on the word-list, *mother*, translates as 'madre,' i.e., with *a*. Much more important is the fact that, although not as striking, it is possible to find Spanish-influenced pronunciations of [ʌ] with a fairly lax and raised variant of [a],[ɐ], as in [sɐŋwɐŋ] 'someone,' and [sɐmbɐði] 'somebody.'[39] Whether these are equally as widespread as the [o] pronunciations is difficult to judge, since they are not as obvious to an English speaker: being unrounded and somewhat farther forward, they appear to coincide better with the phonetic range of [ʌ]. This confusion of Spanish [o] *and* [a] for [ʌ] may partly be due to the AE use of *o* to indicate both phonetic [a][40] and [ʌ], as in the examples quoted above of *mother* and *money*. In my experience as an ESL teacher, certain common spelling confusions of Spanish-speaking students, such as *most* with *must* argue that [o] is definitely confused with [ʌ]. But the same students also fail to distinguish between the spellings of *sang*, *sung* and *song*, suggesting a more general confusion between [a], [o] and [ʌ].[41] It is possible that in Spanish-influenced English, [o]

---

[39] My impressions have been confirmed by Professor William Stewart and Dr. John Roy (personal communications). Dr. Roy further remarks that AE [ʌ] is at an ideal articulatory height for interpretation as both [o] and [a] by Spanish speakers, since it sits on the cusp between these vowels.

[40] For instance (in addition to the *forgat/forgatten* examples cited above), other common ESL-student Spanish-influenced spelling errors are *not* as *nat* and *body* as *bady*.

[41] Support for idea of the Spanish-speaker's imperfect distinction between all three AE vowels is provided by Schnitzer (1997: 77), who notes that the AE phonemes [a], [ʌ] and [ɔ] (as in *cot*, *cut* and *caught*) are all

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and [a] may be in complementary distribution for speakers aiming at [ʌ] and that a variety of factors may decide the choice.[42] In sum, it is possible that [a] pronunciations for *huh* were given, and that the experimenter was unaware of them, due to the closeness of [ɐ] (the [a]-allophone most likely to have been used) to her own [ʌ].

Support for this view is provided by a third piece of evidence. During practice with the word lists and later, while responding to the questionnaire, the Spanish-speaking subjects often complained that the stimuli <a> and <ʌ> sounded confusingly alike, and the experimenter would be asked to clarify the difference between them. At the time, unfortunately, no steps were taken to quantify the very strong impression that this group's <ʌ/a> confusion was widespread. However, on every occasion, it appeared that at least one, more frequently several, Spanish speakers complained about difficulties in distinguishing between <a> and <ʌ>.

Conclusions about the vowels of <ʌ> and <a>. It has been argued that of the L1 possible choices as references for *huh*, the subjects chose [a], in a variant close enough to [ʌ] for English-speaking ears not to register a pronunciation problem. Since the other two

---

liable to be confused by the Spanish speaker. Since these involve both *o* and [ɔ], additional confusion between orthography and pronunciation seems likely.

[42] This was suggested by Professor William Stewart (personal communication). Possible factors are 1) Phonological environment: labial or velar consonantal environments will generate an [o] (*cup, up, buck, punk, mug, gum, bum*, etc.) whereas coronal environments give rise to [a] (*nuts, under, shut, touch, run, justice*, etc.); 2) Conflict between English and Spanish orthographies, further complicated by cognate forms (as in the case of *justice* or *accustomed*); 3) Experience and confidence with English: this may influence those who were [o]-users initially to change to [a] as they feel better able to speak the second language; 4) Sociolectal variation: [o] is felt as more working-class and "Spanish" and [a] more middle-class and "English;" 5) Regional or national variation (such as Dominican vs. Argentinian Spanish-influenced English); 6) True free variation. Clearly, this whole area merits investigation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

possible candidates, [o] (in the lowered variant [o˅]) and [u] (in the lax, lowered variant, [ʊ˅]),[43] have already been excluded as unlikely, besides[ɐ] there is only one other phone which can be considered as a candidate for <ʌ>. This is [a˄], the very lax, raised allophone of [a]. However, [a˄] is found only in intertonic, "very weak" position (Navarro Tomás, 1967: 57, 62), and the experimental instructions did not include this environment and could not have been interpreted to include it.[44] Thus, we conclude that the most likely choice for <ʌ> is [ɐ].

This leaves us with the issue of the most likely realization of <a>. It has been suggested earlier that in being pronounced with an /a/-allophone, the vowel chosen for <ʌ> potentially encroaches on <a>'s territory. Let us make the reasonable assumption that subjects' intentions would be to *make sense* of those questions contrasting <ʌ> and <a>, as they did of, e.g., questions contrasting <i> and <a>. Obviously, they would attempt to keep <ʌ>'s vowel phonetically (and thus, psychologically) as separate as possible from that of <a>. Therefore, the two most polarizable allophones of /a/—rather than two shading into each other (like central [a] and [a˂], or central [a] and [ɐ])—are likely to have been used.[45] For this reason, the tense palatal allophone, [a˂] (which contrasts with

---

[43] For details of these allophones and their distributions, see Colton (190: 35) and Navarro Tomás (1967: 57, 62).

[44] This is clear from the form of the questionnaire items (argued above as possibly leading to primary and secondary accenting on both stimuli, due to pairing in both possible sequences). However, "especially weak position" (Navarro Tomás 1967: 48) is found *between* a primary and a secondary accent, i.e., as the central, unaccented syllable in a string of three; no such interstressed position was available.

[45] For a parallell argument on the division of Japanese /ɯ/ into the polarizable allophones [ɯ̟] and [ɯ˖] (to cover <u> and <ʌ>) see Chapter 9, Section 9.1.1. It is true that since an allophone is no more for the native speaker/listener than a reflex of its phonetic environment, it should be and indistinguishable from its sister-allophones. But clearly, *there are some allophones that, when brought into conscious awareness, are more*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

[ɐ] along three parameters, anteriority, height and laxness) would have been a better choice than [a].

In those questions where <ʌ> and <a> were not being contrasted with each other but with one of the high vowel stimuli, the double usage of /a/ need not have been a problem; in these cases, one variant may have been used for either <ʌ> or <a>.[46] Or the subjects may have consistently used one variant for <ʌ> and another for <a>, even on questions where they were not compared. Finally, evidence for the division of /a/ to cover the pronunciation of both stimuli would be found in <ʌ/a> collapses in the data, especially if occurring both on size and on one (or both) of the vitality dimensions.

Final h. One more question remains to be answered before coming to a conclusion about the pronunciations of *hah* and *huh*: Were their final *h*'s realized in any phonetic form? In Spanish, *h* is never found syllable- or word-finally, but English syllable- and word-final *h* does occur in a limited number of words, with three possible phonetic realizations:

---

hearable than others as differing from the parent phoneme or from others of their sister allophones. For example, one can fairly easily get naïve native English-speaking listeners (such as a class of beginning phonetics students learning the concepts of the phoneme and the allophone) to to notice that such English /t/-allophones as [ʔ] and [D] are different from [t]. On the other hand, it is more difficult to get those same speakers to hear the differences between [t^h], [t^s] and [t^y] and the classic phonemic [t]. With /t/, the suppression of all oral features for [ʔ] or the addition of voicing for [D] are much more *noticeable* changes to the parent phoneme than are the addition of aspiration, affrication or palatalization. Similarly, it is to be expected that in Japanese, fronting or rounding/backing would be felt as very evident changes to the central /ɯ/ phoneme. This is how, it is argued, the Japanese subjects kept the allophonically related vowels of <ʌ> and <u> phonosymbolically separate from each other on 4/6 scales. (I am indebted to Professor William Stewart for reminding me about the non-distinctiveness of allophones and making me clarify this point.)

[46] If so, there are two reasons that the low, tense central variant of /a/ is most likely to have been used for either stimulus. Firstly, Navarro Tomás (1967: 54) characterizes it as the "normal" form of [a]: thus, unless there are evident segmental or stress-based influences on it, pulling it front or back, this would be the 'default' choice. Secondly, in addition to its openness, its centrality would provide an extra feature of contrast with both the frontness of <i>'s vowel *and* the backness of <u>'s.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1) [Ø] consonantally but giving an independent open-syllable short vowel quality, e.g., [a], [ʌ] [ɛ], as in *hah, huh, heh*; 2) [f], as in *laugh, cough, enough*, and also *graph, lymph, hieroglyph*; 3) [Ø], as in *sigh, through, although*. Evidently, for relatively unsophisticated readers, this could produce insecurity about the correct pronunciation of any word-final *h*. Moreover, in the experiment, there were neither written instructions nor oral/aural stimulus modelling to indicate that the final *h*'s were mute, i.e., all of the stimuli should be pronounced as open syllables.

Although the open-syllable short-vowel interpretation was the one intended, it is unlikely that the stimulus spellings were understood this way. There is no reason to assume that any of the Spanish-speaking subjects (apart from # S100, who was not an ESL learner) would have had enough experience with written English to interpret such a specialized use of *h* correctly. The likelihood of final *h*'s being realized as [f] by these subjects is also very small; in the handful of words with this spelling, *h* is always preceded by *g* or *p*, which was not the case with *huh* and *hah*. Moreover, [hVf] pronunciations would have been extremely noticeable and would have been discouraged by the experimenter. Finally, whether subjects recognized *h* as [Ø] and consciously pronounced it as such is clearly impossible to decide; however, since *g* always precedes *h* in spellings indicating [Ø], this seems unlikely.

Thus, the chances of an English-based orthographic interpretation of the final *h*'s in these stimuli are unlikely; however, we cannot rule out Spanish orthographic interference so

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

easily. Since *hah* and *huh* are orthographically closed syllables and end in a character whose interpretation is ambiguous, their spellings could have suggested a checked syllable pronunciation of their vowels, with or without some form of following consonant.[47] One possible realization of the final *h* is as Spanish [x], which as a velar fricative can be seen as equivalent to English glottal fricative [h].[48] However, [hVx] pronunciations for *hah* and *huh* would have been highly noticeable and discouraged by the experimenter, so this is unlikely. A second possibility is that the final *h* was realized as [h], the the word-final (prepausal, preconsonantal) allophone of /s/ [49] Although such pronunciations would be less noticeable, and therefore more likely to escape the experimenter's attention, they too are unlikely. They tend to occur in relaxed, rapid speech, and would not be expected in the phonetic environment of pronouncing separate stimulus syllables, above all AE ones (see arguments above suggesting that each stimulus received main stress). Also, as a stigmatized speech feature, they are unlikley to have been produced in the doubly-formal social setting of an experiment taking place in an ESL classroom.

---

[47] One of the primary sources of Spanish allophonic variation is the contrast between the free and checked nonlow vowels, found respectively in open and closed syllables (see the discussion above, also Colton 1909; Navarro Tomás 1967; Schnitzer 1997). For the non-low vowels, this free-checked contrast generates variations in height and anteriority rather than in duration; however, there are explicit indications in the case of /a/ that its free variants are have a longer duration than the checked ones (Colton 1909: 29-30).

[48] This can be seen in Spanish comic-book renditions of AE *ha-ha* as 'xa-xa.'

[49] There are many descriptions of post-vocalic [s]-aspiration to be found in the literature on Spanish dialect variation (Barrutía and Schwiegler, 1994; Lipski, 1994) since it is considered a distinguishing feature of Andalusian and Latin American coastal varieties of Spanish. Post-vocalic [h] could have been part of the phonetic repertoires of a number of the Spanish speakers, as a variant of post-vocalic [s]: thus the L1 string [-ah], (as in [lah mansanah], *las manzanas*, 'the apples') which was certainly not intended to be evoked by the stimulus spelling hah, would have been a possible pronunciation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A third possibility is that the final *h* was realized as [ʔ], i.e., like [h], a non-oral consonant. Neither of the fricatives were felt to be suitable, and so the final position was marked with an orally unspecified consonant, which did not correspond to any phoneme in Spanish or English, but which would still produce a checked syllable. If <a> and <ʌ> were 'closed' in this way by a consonant, especially if the consonant were non-oral, the main effect may have been a reduction in the durations of their vowels.This might well have escaped the experimenter's attention and can therefore be considered as having some likelihood.

The fourth and final possibility is that no actual consonant was produced for the final *h*, but that the preceding vowel was pronounced as checked, i.e., as having lesser duration than those in *hee* and *hoo*. Such a realization suggests that influence was felt purely from the orthographic shapes of the stimuli. Subjects understood the final *h* to be mute, but since (in contrast with *hee* and *hoo*), the vowels of *hah* and *huh* were orthographically single and followed by a consonant, this was marked by comparatively shorter vowel pronunciations for the non-high stimuli. Moreover, possible differences in vowel duration might have been especially highlighted where the questionnaire asked for comparison between one high vowel and one non-high vowel stimulus, i.e., stimuli of both orthographic types. Pronunciations which involved vowel shortening without a following consonant would not have been noticed by the experimenter, and thus may well have occurred.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In sum, either of the third or fourth possible renditions of h could have produced subtle modifications to the pronuciations of the non-high stimuli and so have escaped the experimenter's attention during practice with the word lists. However, since a relation has already been noted between vowel length and phonosymbolic size judgments, this will be taken up again in the analysis of Spanish subjects' evaluations on the size scales. If Spanish subjects did indeed respond to stimulus spellings by making durational differences in the vowels, we would predict collapses of <ʌ/a> on the size dimension, together with evaluations as non-large/small, and, similarly, <u/i> collapses, with evaluations as large/non-small.

Summary of most likely stimulus pronunciations. It is expected that all four stimuli were pronounced with an initial [h], and for <i> and <u>, the higher allophones of Spanish [i] and [u] were the most likely L1 models. These may have been somewhat diphthongized, due to the influence of English phonology, or lengthened, due either to English phonological influence[50] or to Spanish orthographic influence (i.e., in contrast with spellings of the non-high stimuli), or both. Two allophones of /a/ are considered the most likely L1 models for <a> and <ʌ>. These are, respectively, the low, tense palatal [aˇ] and the somewhat higher, lax backed [ɐ]. It is also possible that the non-high stimuli had shortened vowels, due to Spanish orthographic influence (interpretation of final *h* either as [ʔ] or as a checking influence on the preceding vowel).

[50] This would have meant an interpretation of English diphthongization via Spanish phonology.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Dialectal variation: was this a problem? At this point, a final question remains: Could the different national varieties of Spanish spoken by these subjects have produced variations in the pronunciation of the stimulus vowels? The following breakdown of the hundred Spanish-speaking subjects by country of origin indicates considerable heterogeneity:

Table 10.1: Nationalities of the Spanish-speaking subjects

Colombia	41
Ecuador	16
Dominican Republic	9
Peru	6
U.S.	5
Mexico	4
Venezuela	3
Spain	2
Chile	2
Costa Rica	2
Puerto Rico	2
Nicaragua	1
El Salvador	1
Honduras	1
Cuba	1
Panama	1
Guatemala	1
Bolivia	1
Argentina	1

However, by far the most important dialectal difference is not indicated by Table 10.1. Lipski (1994: 7-8), in discussing the different classificatory approaches to Latin American Spanish dialects, proposes that the crucial dialectal difference is not so much between nationalities but between upland and coastal varieties, sometimes occurring within a nationality.[51] Vowel reduction is characteristic of upland Spanish, while various

[51] I am indebted to Dr. Ricardo Otheguy for a preliminary explanation of the crucial highlands vs. coastal dialect division in Latin American Spanish, and for this reference.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

consonantal phenomena (such as [s]-aspiration or deletion, velarization of word-final [n], and so on) characterize coastal Spanish. In this experiment, among the first seven personal background questions on the questionnaire, the subjects were not asked to state their cities or towns of origin in addition to their nationalities. Thus, for example, the largest single nationality group (Colombians) may well have been two linguistic groups in terms of the Spanish dialects they spoke—but the number of each was unknown.[52] And the same is true of the other Latin American nationality groups.

This appears to pose an extra—and insoluble—problem for our understanding of how subjects pronounced both the four stimuli and their L1 equivalent vowels. But it should also be noted that vowels are not considered the main focus of dialect variation in Spanish. Navarro Tomás, in his *Estudios de fonología española* (1946: 31) asserts that as one looks at different national varieties of Spanish, one finds that the vowels are much less subject to variation than are the consonants.[53] In his study *Spanish in Puerto Rico* (1948), Navarro Tomás describes some possible types of allophonic variation which do

[52] However, I am informed (Dr. John Roy, personal communication) that, as CUNY students, the majority—if not all—of the Colombians are likely to have come from the more prosperous upland regions of the country.

[53] "The following observations [on Castillian vowels] are not applicable only to the Spanish vowels as considered in the pronunciation belonging to Castille... [They] correspond in greater part to the general character which these sounds display in all those countries where Spanish is spoken. *In the phonetic structure of the language, the vowels are a more stable and uniform element than the consonants. The differences of pronunciation between the Spanish of some countries and[that of] others affect consonants more than vowels.* However, a rigorous analysis of the vowels, especially if based on Castillian speech, cannot attribute universal value within the Spanish (speaking) world to all the details which can be observed in the pronunciation of these sounds" (Navarro Tomás 1946: 31 emphasis added).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

not necessarily correspond to Castillian phonological patterning,[54] but the basic specifications of Puerto Rican Spanish vowels are not described as very different from those of Castillian.

Two points can be made here. The first is that in the preceding discussion, the most likely pronunciations have been established for the stimulus vowels, and so far there seems to be no dialectal evidence to challenge them. The second is that even if some dialectal variation had modified the stimulus pronunciations of certain subjects compared to others, the phonological—systematic—relations between the vowels would have remained constant for all subjects. Phonosymbolic attributions to basic contrasts (such as height, anteriority, rounding, duration) were what were being tested in the experiment, and there is no evidence that these contrasts would have been affected. For example, even though the Puerto Rican high vowels are articulatorily somewhat closer to [a] than their Castillian counterparts (Navarro Tomás 1948: 44, 46), the phonological relation between [i] and [u] on the one hand and [a] on the other remains unchanged from one dialect to the other. Thus, if for the two participating subjects from Puerto Rico, [i] and [u] represent the highest vowels in their repertoire, any symbolism which might be triggered by vowel height (even though embodied by a lesser articulatory difference) should still be triggered by their high vowels, as much as it would be for the two subjects from Spain.[55]

---

[54] For example, the palatal variant of [a] tends to be more widespread in Puerto Rican than in Castillian Spanish, and while Puerto Rican [a] has the "same timber as in normal Spanish," [i] and [u] are "somewhat more open than in Castillian" (Navarro Tomás 1948: 44, 46).

[55] This contrasts with cases where contrasts in the vowel systems of two dialects are actually different in some respects, for example, in the standard dialects of British and American English. Thus, Standard Southern English has an extra low back rounded vowel not found in Standard American English, while vowel rhotacization is a general feature of Standard American English not present in SSE. Consequently,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Most of the allophonic modifications of Puerto Rican Spanish vowels described by Navarro Tomás (1948) are not considered significant enough to be included in Lipski's (1994) descriptions of dialect variation. Surveying these accounts, it can be seen that almost all of the phonological dialect differences concern consonants and that dialectal variation in the quality of Spanish vowels is negligible, whether Lipski is discussing the coastal *or* highland variety of the speech of a nationality.[56] The same point is strongly made by Mendez (1982: 191) in an acoustic study of the point vowels in English and Spanish,[57] and by default in Barrutía and Schweigler (1994: 239) in the single mention of vowel reduction and devoicing. Otherwise, their focus is entirely on variation in consonants.

---

the vowels of 'calm' and 'cart' (underlyingly /kam/ and kart/) have different vowel qualities in SAE ([ka:m] and [kart]) but the same in SSE ([ka:m] and [ka:t]). If the pattern of oppositions in the two systems is dissimilar, we would expect this to be reflected in some of their vowel symbolisms.

[56] For the largest group of Spanish-speaking subjects, the Colombians, it appears that neither their highland nor coastal varieties have any notable dialect features specific to vowels. (Lipski 1994: 209-213). Ecuador, with the second-largest subject group, is divided by Lipski into three dialect regions (Lipski 1994: 247-249), but in descriptions of all three, only a single vowel feature is noted, and one not likely to have affected stimulus pronunciation. (This is for the central highlands variety: unstressed vowels are reduced, devoiced and shortened "in contact with /s/, preferentially affecting /e/ followed by /o/" (Lipski 1994: 248). The mid vowels [e] and [o] were not among the four tested in this experiment (nor was [s] the initial stimulus consonant. Moreover, it has been established that the pronunciation of the stimuli involved some form of stress on the vowels.)

The third-largest group, the Dominicans, have no noted dialect peculiarities characterizing their vowels (Lipski 1994: 209-213), although I have personally observed (and had confirmed by others) utterance-final vowel devoicing and raising present in Dominicans' speech: see below. Peruvian Andean highland speakers (possible members of the fourth-largest group) have "extreme" vowel reduction only under the same phonetic conditions as for Ecuadorian central highlanders (Lipski 1994: 320) Finally, we see in the index (Lipski 1994: 42—426) that certain consonant phonemes, such as /s/, /rr/, /l/ and /r/, have lengthy entries while there are *no entries at all* for any of the vowel phonemes.

[57] "[A] characteristic of the Spanish vowel system seems to be its stability. Extremely wide variation is found between English dialects, while Spanish has almost none" (Mendez 1982: 191)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

On the other hand, such an account seems suspiciously simple.[58] For example, having become aware of the presence of vowel devoicing and raising in Dominican Spanish,[59] I specifically looked for instances of raising in Lipski's work, but found only one reference.[60] Yet Navarro Tomás (1948: 48) describes vowel devoicing, raising and lowering in Puerto Rican Spanish, and Florez (1951: 37-80) a vertical instability affecting all Colombian Spanish vowels (mid and low vowel raising and and high and mid vowel lowering).[61] Zamora Vicente (1960: 308) also describes general tendencies of vowel raising and lowering to be found both in Peninsular and Latin American Spanish.

It can be concluded that dialectal vowel variations certainly do exist,[62] yet as features of the various dialects of Spanish, they are apparently not considered as important as allophonic variation in consonants. This suggests that there is, as Navarro Tomás'(1948: 31) remarks indicate, some kind of cross-varietal norm for Spanish [a], [e], [i], [o], [u].

---

[58] Another source of vowel instability should be the less articulatorily anchored nature of vowels compared to that of consonants (Dr Lawrence Raphael, personal communication).

[59] Personal observation (from living in a Dominican neighborhood for many years) alerted me to utterance-final vowel devoicing (noticable with the mid vowels, but not with [a]). There appears very commonly to be raising of utterance-final [o] to [u], as in [sinku] (*cinco*), and to a lesser extent [e] to [i], as in [kinsi] (*quince*), but the front vowel raising effect seems to be much weaker or less obvious.

[60] This was to the Spanish of Costa Rica (Lipski 1982: 223)

[61] Professor Stewart (in a personal communication) has warned me that the articulatory placement of vowels in Carribean Spanish is highly unstable—more counter-evidence to Lipski's point of view. There is a great deal of allophonic variation, some of which is free, some conditioned by speech style. Professor Stewart observes that in self-conscious, meticulous speech the vowels are closer since the mouth is more closed, while singing and emphatic style produces more open mouth and vowels. Would the experimental instructions and the classroom setting have tended to influence the subjects toward more emphatic or more meticulous speech? It has already been argued that meticulous speech was more likely, both because of the setting and the unusual nature of the experimental task.

[62] This general point is made and then argued in convincing detail by Del Carmen Mcnair (1996) whose acoustic vowel study of two Panamanian Spanish dialects indicates both spectral and durational differences between the varieties.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Thus, the only remaining doubt is whether the phonetic environment of the stimulus vowels could have provoked dialect-related gross allophonic variation, such as devoicing, lowering, raising or extreme contraction; it appears that the answer is no.

Firstly, it can be argued that the phonetic environments in which the stimulus vowels were uttered (specifically those determining stress) were not suitable for these types of variation. Secondly, the targets were not Spanish but AE vowels, the experimental setting was entirely in English and the procedures (including a considerable amount of pronunciation in unison, with experimenter feedback) pressured subjects to produce a uniform output. And thirdly, noticeable alterations to vowel qualities (such as devoicing and raising) were not witnessed by the experimenter. This is not to say that while individually answering the questionnaire that some dialectally-conditioned variation could not have crept in (if, for example, a subject started to hurry), but there seems no reason to consider it as a major influence on the quality of the stimulus vowels produced. Therefore, it is not expected to have affected the stimulus pronunciations of these speakers.

10.1.2. Acoustic background

Although Navarro Tomás (1967: 35) has claimed "As yet we have no definite data on the acoustic analysis of the timbre of Spanish vowels," such data do exist, even (Skelton 1950) predating this remark. However, as with Japanese, a large-scale acoustic study of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Spanish vowels, equivalent to Peterson and Barney's (1952) for English, does not yet appear to have been made. Of the smaller Spanish studies located for this survey, three do not provide formant values, but their general observations on vowel placement will be included. Where formant values are given, the approach will be to compare several data sets,[63] as was done in the previous chapter for Japanese; AE male and female values from Peterson and Barney (1952) will then be provided for interlinguistic comparison. Finally, while there is a surprising lack of material on intrinsic F_0 for Spanish,[64] two F_0 studies have been located for the present survey, together with some data on vowel duration: these will be discussed.[65] No data have been located for the intrinsic intensity of Spanish vowels.

Studies of interest which do not give formant values. Guirao and Borzone de Manrique (1975), made a spectrograhic analysis of recordings of isolated vowels as uttered by seven male and seven female Argentinian Spanish speakers. They argue for the importance of the comparative intensity of the formants in a vowel as its main identifier;

[63] In interpreting Spanish vowel formant values, the dialect variable will be borne in mind. These studies used informants speaking different national varieties of Spanish, possibly with some differences in the pronunciation of the vowels, and this may well account for some of the incidental differences in formant values where they are quoted. Even so, such differences are not expected to have affected the Spanish subjects' phonosymbolic judgments. The reader is reminded of the point made in the previous section on dialectal variation: however the basic phonological oppositions are rendered from dialect to dialect, they are expected to remain constant in all varieties of Spanish

[64] For example, in a relatively recent survey of data on F_0 which covered 31 languages including French and Italian (Whalen and Levitt 1995) Spanish was notably absent.

[65] For some of the F_0 and all of the durational data, I am vey much obliged to Dr. Winifred Strange (of the CUNY Graduate Center's Speech and Hearing Department) who very kindly made the Master's thesis of her student, Melissa del Carmen McNair, available to me.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with [i] and [e], the hearer is focused toward F_2 and F_3, with [a], the focus is on F_2, and with [o] and [u], on F_1 and F_2 (Guirao and Borzone de Manrique 1975: 22).[66]

Comparing the areas of Argentinian Spanish and the cardinal vowels on a scatterplot diagram, these authors agree with NavarroTomás' (1948, 1966) accounts of Puerto Rican and Peninsular Spanish.[67] Comparing the acoustic areas of Argentinian vowels with those of American English, as given in Peterson and Barney's (1952) scatterplot, they note that Spanish [i] and [u] fall well within the areas of the AE high vowels. Spanish [a] falls between the AE vowels of *hud* [hʌd] and *hod* [had], while "Spanish [o] is similar to the vowel sound in *hawed* [hɔd]" (Guirao and Borzone de Manrique 1975: 25). Thus, they support the observations given in the previous section that AE [ʌ] may be confused by the Spanish speaker with AE [a], and that orthographic AE *o* as in *front* or *mother* could well be interpreted as Spanish [o] and thus pronounced, to the ears of American English speakers, with a vowel resembling AE [ɔ].

Quilis (1981: 174-175) briefly reports on the results of a study using as informants sixteen men and six women of mixed Latin American and Castillian backgrounds. Recordings were made of the five vowels "in normal phonetic position."[68] In his

---

[66] Since Guriao and Borzone de Manrique do not include F_0 in their study, these conclusions are somewhat different from those about the importance of locus (see Chapter 7, Sections 7.2.1 and 7.2.2).

[67] They observe that Spanish [u] is considerably lower than cardinal [u] (CV 8) located midway between CV's 8 and 7. Spanish [i] "overlaps partly" with cardinal [i] (CV 1): again it is lower, but not as low as Spanish [u]. Spanish [a] "shares a small area with (cardinal [a]) CV 4" but tends to be articulated farther back (Guirao and Borzone de Manrique 1975: 22).

[68] Vowels were in a voiceless labial frame: [pVpa].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

comparison of the Spanish vowels with the cardinal vowels, Quilis' findings do not entirely agree with those of the Argentinian researchers quoted above; [i] and [u] were found to be somewhat farther back, and [a] also farther back and higher (Quilis 1981: 175).

Romero (1988) investigated the perception and classification of fifteen cardinal and three English vowels by thirty-two Castillian Spanish-speaking students.[69] Surprisingly, it was observed that the close front vowel, CV 1 (cardinal [i]) had only a 65.7% identification with Spanish [i], being identified with Spanish [u] for 21.8%, and with Spanish [e] on 12.5% of the responses. Identification of [ʌ] was fairly evenly divided in thirds between Spanish [e] at 31.2%, Spanish [a], also at 31.2%, and Spanish [o] at 37.6% of responses. Both CV's 4 and 5 were identified 100% with Spanish [a] (Romero 1988: 188). Romero's subjects' confusion of [i] with Spanish [u] suggests that around one-fifth of them may have had auditory problems.[70] His results associating cardinal [i] and [ʌ] with Spanish [e] are surprising, but the identification of [ʌ] with both Spanish [o] and Spanish [a] bear out the earlier observations about how [ʌ] is perceived and pronounced by Spanish speakers. In Romero's study, Spanish [a] emerges as the vowel with the strongest identity, both in

[69] Eighteen non-Spanish vowels were recorded and played to the subjects, who were asked to identify all of the sounds they heard with one or other of the five Spanish vowels. Spectrograms were then made both of the stimulus and Spanish vowels, and Romero calculated the ranges of values of F_1 and F_2 within which a stimulus vowel was identified with one of the Spanish vowels.

[70] In other words, they did not register the high-frequency components of [i] (F_2 and F_3), and so focused on the lower F_1, which was at an equivalent height for both [i] and [u].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

terms of its undivided association with the low mid and back cardinal vowels and also of the narrowness and precision of its identificatory formant values.[71]

Comparison of Spanish formant values drawn from three sources. Mendez (1982) made a spectrographic study comparing the Puerto Rican Spanish point vowels with their (Midwestern) AE equivalents, using three informants from each language (sex unspecified: from the formant values obtained, presumably all male).[72] Morimoto (1988) used five male and five female informants, speakers of Peninsular Spanish, and compared their vowels with those of the same number of Japanese-speaking males and females.[73] Del Carmen McNair (1996) analyzed the five vowels of two dialect groups, each of six male speakers of Panamanian Spanish, from Panama City and from David (rural dialect).[74] Despite its large number (twenty) of informants, values from an earlier study (Skelton 1950) have not been included, because they are so unusually high as to indicate possible error, and because Skelton's method of arriving at his averaged formants is questionnable.[75]

---

[71] For 80% of subjects, Romero found [a] was unique among the five vowels in having single identificatory values (rather than a range of values) for both F_1 and F_2 (Romero 1988: 201).

[72] All vowels were uttered in isolation, and acoustic measurements were taken at the midpoint of each sample.

[73] The vowels analyzed were uttered in bisyllabic words under four conditions: as stressed first nucleus, stressed second nucleus, unstressed first nucleus and unstressed second nucleus. Morimoto does not specify from which point in the recording his analyzed samples were taken.

[74] Recordings were of the same frame sentence which contained stressed specimen vowels in bisyllabic [CVCa] words with normal Spanish accenting: with stress on the penultimate (i.e., first) nucleus. Formant values are based on an average of midpoint samples.

[75] Thus, Skelton's (male) values for the F_1's of both [i] and [u] are 504 Hz, which is in striking contrast to those of the other F_1 values obtained for the male Spanish high vowels (see below, Table 10.2). It should be born in mind that Skelton's study predates that of Peterson and Barney, i.e., that it is very early in the history of this type of acoustic investigation. Moreover, Skelton's values are based on on the averaging of acoustic measurements of each vowel under all of its possible phonetically conditioning environments, i.e.,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Due to the smallness of the number of informants and the diversity of Spanish dialects represented, the survey which follows should be considered as giving no more than a reasonable indication of the placement of Spanish vowels and their relation to those of American English. Since only one group of female Spanish formant values were located (those of Morimoto 1988), these are presented later, in interlinguistic comparison with the equivalent female values from Peterson and Barney (1952). As in Chapters 8 and 9, minimal discrepancies between the values of two formants have been fixed as indicating true articulatory divergence.[76] Values for the Spanish mid vowels have been included as an aid in comparisons along both articulatory dimensions.

Table 10.2: Comparison of male F_1 values (Hz), drawn from three Spanish studies.

	Mendez *P. R.*	Morimoto *Peninsular*	del C McN *Pan. City*	del CMcN *David*
[a]	756	750	618	582
[e]	-----	450	439	443
[i]	350	290	321	308
[o]	-----	530	439	446
[u]	336	360	355	339

under all of the contextual conditions of accenting types, word position and contiguous segmental influence (Skelton 1950: 4-5). But it would appear from the discussion above (in Section 10.1.1) of the accenting of Spanish vowels, that "very weak" (Navarro Tomás, 1967) intertonic position fundamentally changes the nature of non-low vowels (shortening, laxing, lowering, centralizing, etc.), and that similar changes apply to final unstressed [a]. It is therefore possible that inclusion of these weak vowel tokens skewed Skelton's analysis and the formant values which he obtained.

[76] A difference of more than 75 Hz between the values of two F_1's is considered to show vertical divergence: an increase of +75 Hz or over indicates comparatively more jaw-openness/tongue height, while the equivalent decrease (—75 Hz. or under) indicates a comparatively closer jaw or lower tongue position. Similarly, a discrepancy of more than 125 Hz. between the values of two F_2's is considered as indicating horizontal divergence; an increase of +125 Hz. or over indicates a more anterior articulation, and the equivalent decrease (—125 Hz. or under), a more backed (or more rounded) vowel.Guidelines here were provided by Dr. Lawrence Raphael (personal communication).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Adapted from Mendez (1982: 193), Morimoto (1988: 185) and del Carmen McNair (1996: 57).

Table 10.3: Comparison of male F_2 values (Hz), drawn from three Spanish studies.

	Mendez *P. R.*	Morimoto *Peninsular*	del C McN *Pan. City*	del C McN *David*
[a]	1178	1460	1399	1387
[e]	-----	2160	1949	1862
[i]	2478	2480	2254	2186
[o]	-----	950	881	904
[u]	751	860	803	837

Adapted from Mendez (1982: 193), Morimoto (1988: 185) and del Carmen McNair (1996: 57).

Inspection of Tables 10.2 and 10.3 show similar F_1 values for the [a]'s of Mendez and Morimoto, indicating vowels of equal openness/low tongue position; the lower F_1's for both of del Carmen McNair's [a]'s suggest closer articulations, with that for David speakers even less open than that for Panama City.[77] The F_2 values for [a] show Mendez' vowel farther back than the other three; Morimoto's is the most anterior, but F_2 values for both of del Carmen McNair's [a]'s are within Morimoto's range. The F_2 discrepancy between Mendez' [a] and the others' bears out earlier observations (in this, and Chapters 7, 8 and 9) on the tendency of [a] to vary horizontally, whatever the language.

Comparison of F_1 values for [i] indicates that Morimoto's Peninsular Spanish speakers may show a slightly closer/higher articulation[78] than the vowels of the other three groups,

[77] We note that all of the three F_1 values given for [e] indicate equal height, but those for [o] suggest higher articulations for the Panamanian speakers than for the Castillians.

[78] However, the value discrepancy here (of 60 Hz.) is below the difference established as significant for F_1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which all agree, indicating vowels of equivalent height. Mendez and Morimoto's F_2 values for [i] are almost identical, and indicate a very forward placement; the two Panamanian groups are somewhat less anterior and similar to each other's. For [u], similar F_1 values for all four groups indicate equivalently high vowels;[79] their F_2's are also similar, indicating an equivalent degree of backness/rounding for the [u]'s of all four groups.[80]

Spanish formant values compared with those for American English. First, we present data from female speakers.

Table 10.4: Comparison of female F_1 values (Hz), drawn from one Spanish and one American English study

Sp.	Morimoto *Peninsular*	AE	Peterson & Barney
[a]	990	[a]	850
[e]	480	[ɛ]	610
[i]	350	[iy]	310
[o]	600	[ɔ]	590
[u]	410	[uw]	370
		[ʌ]	760

Adapted from Morimoto (1988: 185) and Peterson and Barney (1952: 183)

[79] Or they may also indicate equivalent tightness of lip rounding (Dr Lawrence Raphael, personal communication).

[80] For the Peninsular Spanish and David speakers, higher F_2 values for [o] indicate that it is forward of, or less rounded than [u]—the opposite of the Japanese case—while F_2 values for the Panama City speakers' [o] and [u] indicate vowels of equal backness/rounding.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 10.5: Comparison of female F_2 values (Hz), drawn from one Spanish and one American English study.

Sp.	Morimoto *Peninsular*	AE	Peterson & Barney
[a]	1680	[a]	1220
[e]	2570	[ɛ]	2330
[i]	2890	[iy]	2790
[o]	1070	[ɔ]	920
[u]	940	[uw]	950
		[ʌ]	1400

Adapted from Morimoto (1988: 185) and Peterson and Barney (1952: 183)

Inspection of Tables 10.4 and 10.5 shows similar F_1 values for Peninsular Spanish and AE [a]'s, indicating equivalent openness/low tongue articulation for these vowels. However, when F_2 values are compared, we see a much more anterior placement for Spanish than for AE [a]. F_1 values for the [i]'s of both languages suggest equivalent closeness/tongue height; a somewhat higher value for the Spanish than the AE F_2 suggests a possibly more forward articulation for Spanish [i].[81] F_1 values for the [u]'s of both languages indicate the same degree of closeness/tongue height (or tightness of lip rounding), and similar F_2's indicate vowels of equal backness.[82]

[81] However, at 100 Hz. difference between the two F_1's, this is inside the 'similarity' range.

[82] We note that while AE [ɔ] and [uw] have F_2's indicating equal backness, difference in female F_2 values for the Peninsular Spanish vowels indicate [o] as well forward of/less rounded than [u]. This F_2 difference in values (and hence, horizontal placement) is not so marked for male Peninsular Spanish [o] and [u] (see Table 10.7, below).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Next, we present male formant data. The AE vowels from Mendez' (1982) study have been included as a point of comparison with those of Peterson and Barney.

Table 10.6: Comparison of male F_1 values (Hz), drawn from three Spanish and one American English study.

Sp.	Mendez *P. R.*	Morimoto *Peninsular*	del C McN *Pan. City*	del CMcN *David*	AE	Peterson & Barney	Mendez
[a]	756	750	618	582	[a]	730	765
[e]	-----	450	439	443	[ɛ]	530	-----
[i]	350	290	321	308	[iy]	270	306
[o]	-----	530	439	446	[ɔ]	570	-----
[u]	336	360	355	339	[uw]	300	364
					[ʌ]	640	-----

Adapted from Mendez 1982: 193), Morimoto (1988: 185), del Carmen McNair (1996: 57) and Peterson and Barney (1952: 183)

Table 10.7: Comparison of male F_2 values (Hz), drawn from four Spanish and two American English studies.

Sp.	Mendez *P. R.*	Morimoto *Peninsular*	del C McN *Pan. City*	del C McN *David*	AE	Peterson & Barney	Mendez
[a]	1178	1460	1399	1387	[a]	1090	1211
[e]	-----	2160	1949	1862	[ɛ]	1840	-----
[i]	2478	2480	2254	2186	[iy]	2290	2410
[o]	-----	950	881	904	[ɔ]	840	-----
[u]	751	860	803	837	[uw]	870	968
					[ʌ]	1190	-----

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Adapted from Mendez 1982: 193), Morimoto (1988: 185), del Carmen McNair (1996: 57) and Peterson and Barney (1952: 183)

Inspection of F_1 values for AE [a] indicates similar degrees of openness/low tongue position for Peterson and Barney's and Mendez' vowels.[83] AE values are in agreement with the more open vowels of Mendez and Morimoto, but contrast with the closer [a]'s of del Carmen McNair, particularly that of the David speakers (whose [a] appears of equivalent height to the mid-low AE vowel, [ɔ]). A range of F_2 values is seen for [a], indicating that Peterson and Barney's is the most back, Mendez' AE vowel somewhat more anterior,[84] and Mendez' Puerto Rican [a] between the two in placement. The F_2's of Morimoto's and both del Carmen McNair's groups indicate a considerably more anterior placement for all three [a]'s. Thus, the Spanish F_2 values surveyed here (apart from those of Mendez) suggest a more front articulation for Spanish [a] than for its American English counterpart. This agrees with the articulatory descriptions of Sorzano Jorrín (1928) and Navarro Tomás (1967), who characterize [a] as a central vowel.

Surveying F_1 values for [i], we note that Mendez' Spanish vowel has a definitely lower placement than Peterson and Barney's [iy], but that their [iy] agrees in value with Mendez' AE vowel and with the other three Spanish [i]'s, indicating equivalent closeness/tongue height. Comparison of the AE F_2 values for [i] indicates that Mendez'

---

[83] All of the F_1 values for Peterson and Barney's and Mendez' AE vowels are in agreement, with those for [uw] showing the most divergence. Mendez' AE F_2 values (though within the 125 Hz. limit) indicate a possible more anterior placement than Peterson and Barney's for all three vowels, with values for [uw] the least divergent.

[84] *We note that, at 121 Hz., the discrepancy between the two AE [a] F_2's is on the cusp of difference.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

vowel is somewhat more anterior than Peterson and Barney's,[85] and overall that there are two cross-linguistic horizontal placements for [i]. Mendez' AE and Spanish vowels and Morimoto's are more anterior, while Peterson and Barney's, and both del Carmen McNair's are somewhat more central.

F_1 values for [u] indicate fairly equivalent height for all six vowels. Peterson and Barney's F_2 for [uw] is within the range of those of all of the Spanish [u]'s (including Mendez', placed slightly farther back than the others). However, the F_2 differences between Mendez' AE [uw] and the Spanish [u]'s of Mendez and del Carmen McNair indicate that Mendez' AE vowel is anterior to all of the Spanish vowels except Morimoto's.

To sum up, among the four dialects of Spanish represented (by male speakers), we see evidence of articulatory differences along both vertical and horizontal articulatory dimensions for [a] and [i]; in contrast, [u] appears more stable. Comparing Spanish with AE vowels, we see horizontal interlinguistic differences between the [a]'s of both male and female speakers. For [i]/[iy], we see possible (for females) and definite (for males) interlinguistic differences in horizontal placement, while [u]/[uw] appears generally more stable. Despite some divergences, the formant values of the point vowels of both languages are generally within each other's ranges, and this suggests more interlinguistic similarities than differences in their placements.

[85] The discrepancy between the two AE F_2's of [i] is 120 Hz, i.e., a cusp value.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Locus. Only del Carmen McNair's data cover the full range from F_0 through F_3; her values are given below:

Table 10.8: Male formant values for F_0 through F_3 (Hz) of two dialects of Panamanian Spanish

	Panama City				***David***			
	F_0	F_1	F_2	F_3	F_0	F_1	F_2	F_3
[a]	133.6	618	1399	2305	113.6	582	1387	2211
[e]	137.0	439	1949	2459	115.5	443	1862	2381
[i]	**140.4**	**321**	*2254*	*2714*	**118.1**	**308**	*2186*	*2603*
[o]	137.4	439	881	2376	117.0	446	904	2290
[u]	**141.1**	**355**	803	2314	**118.9**	**339**	837	2333

Adapted from del Carmen McNair (1996: 57, 65)

Bolded values for F_0 and F_1 indicate that in both dialects there are low locii of concentration of acoustic energy for [u], and [i]; italicized underlined values for F_2 and F_3 indicate a high locus for [i]. Absence of bolding for [a] indicates lack of a locus in the F_1—F_2 range, due to [a]'s rather low F_1 and relatively high F_2 values, together suggesting more central placements for these vowels. It has already been observed that more front articulations of [a] will result in a wider F_1—F_2 separation, and thus an [ʌ]-like diffusion of acoustic energy, rather than a concentration in the middle range.[86] Since the Puerto Rican [a] reported by Mendez is the only one with a relatively low F_2 value, this is the only Spanish vowel which can be claimed to have a middle locus, as illustrated below. Bolded values indicate middle locii for both AE and Spanish [a]'s

[86] For discussion of this point, see Chapter 9, Section 9.1.2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 10.9: Puerto Rican Spanish and AE male F_1 and F_2 values (Hz).

	Spanish		*AE*	
	F_1	F_2	F_1	F_2
[a]/[a]	**756**	**1178**	**765**	**1211**
[i]/[iy]	350	2478	306	2410
[u]/[uw]	336	751	364	968

(Adapted from Mendez 1982: 193)

Though Mendez and Morimoto do not provide the full range of F_0—F_3 values, there are no reasons to believe why the high vowels measured in their studies would not have had acoustic-energy concentrations similar to those of del Carmen McNair's; thus, we would expect the presence of locii for <i> and <u>, as pronounced by the Spanish-speaking subjects. However, it is equally likely that a locus of acoustic energy concentration would *not* have been present for this group's <a> any more than for their <ʌ>. These stimuli are argued to have been realized as the allophones of Spanish /a/, [a˂] and [ɐ], respectively the low, tense, more front, and raised, lax, more back variants (see above, Section 10.1.1). The front placement of [a˂] would serve to increase its F_2, while the raising of [ɐ] would diminish its F_1. Either factor would have had the effect of widening the acoustic gap between the first two formants, and thus producing a diffusion of of acoustic energy in the middle range.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Intrinsic duration. One study alone has been located giving durational data: Del Carmen McNair's (1996),[87] which, as described above, used two small dialect groups of six subjects each. Given the smallness of her sample and the particularity of the dialects studied, what follows can be considered therefore as no more than a likely indication of the durations of Spanish vowels. We present her findings below, with the mid vowels included for comparison:

Table 10.10: Male durational values (ms) for the vowels of two dialects of Panamanian Spanish

	Panama City	*David*
[a]	131.3	114.3
[e]	123.3	107.3
[i]	110.7	87.3
[o]	127.0	108.6
[u]	107.7	93.0

Adapted from del Carmen McNair (1996: 54)

Two features are clearly seen in Table 10.10. Firstly there is an overall dialect-based difference in durations, with Panama City speakers having longer vowels. Secondly, consonant placement apart[88] (which is irrelevant for this experiment) one feature alone, openness/tongue height, seems to determine vowel duration (Del Carmen McNair 1996: 51). For both groups, [a] is the longest vowel, the mid vowels have similar, lesser, durational values, while the high vowels are the shortest, with similar values. Thus, the

[87] Durations of the vowels were measured in three consonantal contexts :[bVba]. [dVda], [gVga], and values for the three contexts were then averaged.

[88] Anteriority of flanking consonants was also found to have a durational effect (del Carmen McNair 1996: 69).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

findings indicate a hierarchy of three steps, and while vowel durations may vary with dialect, the overall patterning is not affected.[89] We note that for the purposes of this experiment, since the mid vowels were not included in the investigation, durational ranking would translate as a binary opposition (i.e., of long <a> vs. short <u/i>).

Intrinsic F_0. Two studies provide us with data here. Mateo, (1988) investigating intrinsic F_0 in Spanish, used fifteen male and fifteen female Castillian speakers as informants. A graph is given of accented and unaccented vowels[90] (Mateo 1988: 163) in both female and male speech: as the accented condition corresponds to that for the stimulus vowels were uttered in the present experiment, it is of obvious relevance. For both sexes, the high vowels have the highest F_0's, with very similar values for both [i] and [u] (around 226 Hz for females and 154 Hz for males). The mid vowels have lower F_0's and are also close in values (around 212 Hz for females and 145 Hz for males), while [a] has the lowest F_0 (around 205 Hz for females and 143 Hz for males). As Mateo (1988: 164) notes, this suggests a direct connection between F_0 and vowel height.

In her study of Panamanian Spanish, Del Carmen McNair (1996), presents the same F_0 picture as Mateo, further supporting the hypothesized link between fundamental frequency and vowel height. Her findings are given below:

---

[89] This maintenance of patterning despite interdialectal differences has been argued as central to the issue of dialect-based articulatory variation (as discussed above, in Section 9.1.1) and as the reason for expecting consistency of phonosymbolisms among speakers of different varieties of Spanish.

[90] *Vowels were presented, accented and unaccented, as uttered in sentences containing the syllable [pVp].*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 10.11: Male values for F_0 (Hz) of two dialects of Panamanian Spanish

	PC	*David*
[a]	133.6	113.6
[e]	137.0	115.5
[i]	140.4	118.1
[o]	137.4	117.0
[u]	141.1	118.9

Adapted from del Carmen McNair (1996: 65)

While values are overall lower for the David speakers than they are for those from Panama City, the same relation between F_0's can be seen for both groups, with the high vowels having the highest, very similar, F_0 values, the mid vowels somewhat lower F_0's, also close in value, while [a] has the lowest F_0.

Taken together, Mateo's and Del Carmen McNair's findings suggest that for the F_0's of the stimulus vowels in this experiment, we should expect an opposition between <i/u> and <ʌ/a> (since both are argued to derive from /a/).[91] Clearly, this is very similar to the opposition seen for durational values. The agreement between Spanish intrinsic duration and intrinsic F_0 data suggests that both derive from the basic close/open contrast, and this is likely to be phonosymbolically important for Spanish speakers. Collapses of <i/u> and <ʌ/a> may therefore be expected, especially on the size dimension.

[91] But it is also possible that <ʌ>, argued to derive from [ɐ], would count as a mid-height vowel, and therefor occupy a middle F_0 ranking.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

10.2. Analysis of the Spanish data

Figure 10.1 below presents in summary form the stimulus rankings on the six semantic scales, derived from the responses of the Spanish speakers. It corresponds to Figures 7.2, 8.2 and 9.1, summarizing, respectively, the repsonses of the AE, Korean and Japanese subjects. Like the earlier figures, it is based on indices and is accurate in scale; rankings are integrated with consensus data, with only the more unusual cases of non-differentiation being indicated.

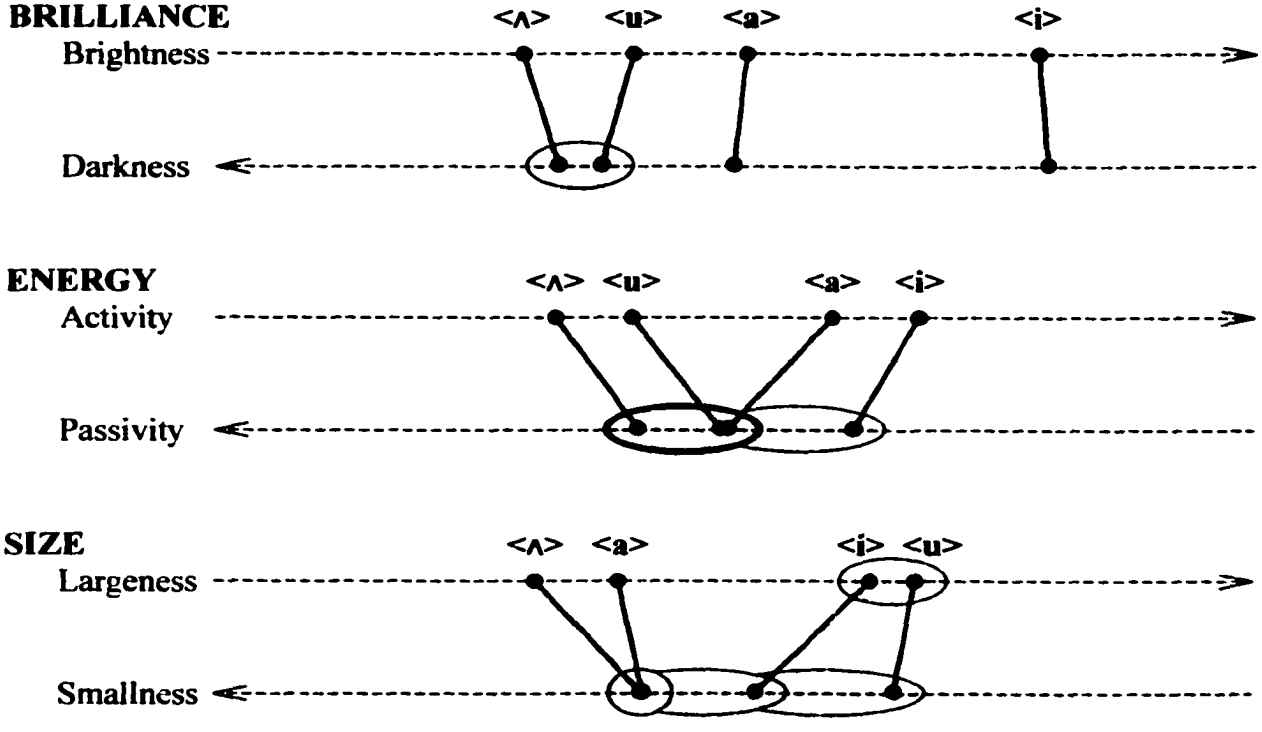


Figure 10.1: Data patterns (Spanish speakers). Ranking index values and consensus data for salient and non-salient sub-scales of three semantic dimensions. Ellipses surround stimulus pairs not differing by statistical test, p>.05; bolding indicates a special case with no significant contrasts among a set of *three* stimuli.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A first point, on inspecting Figure 10.1, is once again the broad similiarity between the sets of stimulus rankings obtained on the brilliance and energy dimensions. Since all four data sets have now shown (intralinguistic) brilliance-energy ranking agreements, we see unconflicted evidence for the 'vitality' superdimension. Spanish vitality rankings are the same as those of AE; their most obvious disagreement with those of the Asian language speakers is that <i> (rather than <a>) is considered the most vital stimulus. The Spanish size rankings are strikingly different from those of all the other language groups in two respects: the collapse of <u/i> at the largest/least-small endpoints, and the second-least-large/smallest ranking for <a> and its collapse with <ʌ> on smallness. These findings support the hypothesis that influences not controlled for in the experimental design affected the size judgments of Spanish speakers.[92] We also note that the salient size and non-salient brilliance scales resemble each other unusually in having symmetrically similar configurations of two independently-ranked stimuli and one collapsed pair (cf. the chaining on the salient size scales of the other three groups).

While both non-salient size and energy scales show extensive nondifferentiation (suggesting that subjects found evaluations difficult on both), they have different configurations. Energy, with four nondifferentiated pairs, shows a three-way collapse, plus a collapsed pair and indeterminate placement of the shared stimulus, <a>; size, with three nondifferentiated pairs, shows chaining, in which the collapsed <ʌ/a>, <ʌ/i> and <u/i> produce indeterminate placements for <ʌ> and <i>.

[92] These were stimulus orthography ('short' final *hVh* versus 'long' *hVV* spellings) and, L1-based reinterpretation of the tense/diphthongized AE target high vowels as longer than the non-high ones. See above, Section 10.1.1, for full discussion of these issues.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Overall, we see consistent endpoint rankings of <i> (as bright/non-dark, active/non-passive and large/non-small), but placements are shared with <u> on size and <a> on passivity. We also see consistent opposite-endpoint rankings for <ʌ> (as non-bright/dark, non-active/passive and non-large/small): these are independent on the salient scales, but are shared with <u> on darkness, <u/a> on passivity and <a> on smallness.

A second point arising from inspection of Figure 10.1 is the symmetricality of rankings for the salient and non-salient scales.[93] As for the three previous data sets, the non-salient scales have narrower spacing than the salient. Range disparities agree with those shown by the AE, Korean and Japanese groups in some aspects, but differ strikingly from them in others, as the following table illustrates.

Table 10.12: American English, Korean, Japanese and Spanish response range and discrepancy values for salient and non-salient scales of three semantic dimensions

	AE			Korean			Japanese			Spanish		
dim.	*sal.*	*disc.*	*n-sal.*	*sal.*	*disc.*	*n-sal.*	*sal.*	*disc.*	*n-sal.*	*sal.*	*disc.*	*n-sal.*
brill.	.710	.096	.614	.530	.115	.415	.591	.029	.562	.497	.025	.472
engy.	.612	**.165**	.447	.541	**.343**	.198	.541	**.198**	.343	.350	**.140**	.210
size	.342	.059	.283	.359	.106	.253	.436	.044	.392	.365	.119	.246

Bolded numbers indicate that in all four data sets the energy dimension consistently shows the largest response-range discrepancy. However, we note that the value of the Spanish range discrepancy for size is almost as large as it is for energy, and that the

[93] There are no ranking asymmetries in the Spanish data.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

salient size scale has a larger range than that of the salient energy scale (both underlined): in this respect, the Spanish size results are different from those of the other three groups.

While brilliance and energy yield similar rankings, we therefore see considerable differences between them in the Spanish data. The ranges for the salient and non-salient brilliance scales are almost equally wide; the salient energy-scale range is considerably reduced in comparison (more so than for any other group) and the range of the non-salient energy scale, further reduced. Thus, Spanish response range values suggest three groupings of ease/difficulty: both brilliance scales were the easiest to make judgments on, while the salient energy and size scales were somewhat more difficult, and most difficult were the non-salient energy and size scales. We conclude that while salient and non-salient size and energy judgments were of roughly comparable difficulty,[94] salient-scale size judgments appear to have been easier for the Spanish speakers than for the other three groups.

In analysis of the other groups' size data, it has been suggested that evaluation according to more than one phonetic criterion (trigger) was responsible for the more complex configurations and smaller response ranges of the size scales. For the Spanish subjects, it would appear that an additional, more unifying criterion may have been acting in the case of the largeness scale. It has been argued above (in Section 10.1.1) that stimulus spellings and L1 reinterpretation of AE tenseness/diphthongization may have divided the stimuli

---

[94] Similar difficulty in responding to the passivity and smallness scales has also been argued for the Korean and Japanese subjects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

into long/non-long, or long/short for these subjects, vowel length being then reinterpreted as largeness,[95] but these factors should have affected salient and non-salient size scales equally. Consequently, it is also worth surveying the results from the translation page of the questionnaire (as was done in interpreting the anomalously large range-discrepancy value for Japanese energy). Since it was already noted that cognate items in the word lists may have presented pronunciation problems for the Spanish subjects, cognate translations (correct or not) have been assigned a separate column in the table presented below.

Table 10.13: Adjectival terms: Spanish translations,[96] cognates and mistranslations.

Adjective	Total trans.	Cognates	Mistrans.
bright	26	—	1
dark	26	—	1
active	29	23 (*activo*)	2
passive	31	23 (*pasivo*)	7
large	33	21	23 (*largo*: 21)
small	32	—	6 (*corto*: 5)

By the time that the adjective page was added to the questionnaire, only twenty-seven Spanish-speaking subjects remained to be interviewed, one of whose responses was disqualified.[97] While these results cover therefore no more than 26% of the subject pool, they strongly suggest for the Spanish speakers that first-language cognate forms

---

[95] For a discussion of the relation between vowel length and phonosymbolic intuitions of size, see Chapter 7, Section 7.2.3.

[96] Translation totals reflect the fact that some subjects gave more than one translation of an adjective.

[97] Subject # S92 misunderstood the task and, instead of translations, gave the pronunciations of the adjectives as supplied by his English-Spanish dictionary. His response was therefore excluded from the total.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

dominated their sense of the meanings of *active*, *passive* and *large* on the questionnaire. For the adjectives *dark* and *bright*, which involved no cognate forms, almost every response was a correct translation, and in the case of 'pasivo' and 'activo', the English and Spanish forms and meanings corresponded, resulting in relatively low numbers of mistranslations. However, following the same strategy with *large* resulted in a mistranslation, since Spanish word 'largo,' means *long*. We have evidence that at least one-fifth of the total of Spanish subjects understood *large* as 'largo,' either exclusively or as one of two or three meanings. [98] Assuming this group is representative, the high proportion of 'largo' mistranslations indicates that the same misunderstanding is likely to have been the case for four out of five of all Spanish subjects.

Table 10.13 also indicates that whereas there were relatively few mistranslations for *small* (for which no cognate Spanish form exists), 5/6 of those were 'corto,' *short*, i.e., the opposite of 'largo.'[99] The asymmetry of numbers of mistranslations for the size terms corresponds nicely to the range/configuration differences between the salient and non-salient size scales. Since this suggests an asymmetry of semantic terms, it must be asked here whether the smallness and largeness results should be interpreted as separate, unrelated measures, i.e., of length vs. smallness. It can be answered no. The translation factor is only one of three (the subjects' first language; the effects of dimension; the effects of salience) which may have been responsible for the differences between the Spanish largeness and smallness results. On the other hand, the equation of

---

[98] Sixteen out of the twentyone subjects offering 'largo' gave it as the only translation of *large*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phonosymbolic size with phonetic duration has been argued as a possible factor in the other subjects groups' evaluations, although not consciously applied as such.

Moreover, non-salience appears overall to blur responses of most scales,[100] and there is no indication that this would not have happened in any case with the Spanish speakers' non-salient size judgments. Interlinguistically surveying the size results (see Table 12 above, and Appendix 4), we note that response range values, hence salient-vs.-non-salient range discrepancies, are almost the same for the Korean as for the Spanish speakers, and that even greater non-differentiation is found on the Japanese than on the Spanish smallness scales. Thus, while the differences between the Spanish subjects' largeness and smallness evaluations may have been partly due to more vs. less conscious application of the same criterion, they cannot be seen as the result of judgments according to two different criteria. Therefore the scales should not be interpreted as two separate measures. Clearly, however, in designing further experiments using scales of this type, scale terms must be screened for 'falso amigo' cognates.

A third point raised by inspection of Figure 10.1 is the importance of stimulus contrasts arising from the pairing of stimuli in the test items. The Spanish data show no clusters of collapses (unlike the Korean and Japanese cases), nor do they show smaller groupings of three or four collapses (unlike the AE data). On the other hand, the distribution of non-significance is more dispersed than for the other three groups, since all stimulus pairs

---

[99] In every case, these were where the subject had also given 'largo' for *large*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

show at least one collapse. [101] For <u/a> and <a/i> there is a collapse on passivity, and for <ʌ/i>, on smallness. An interaction between dimension and pairing is shown for <u/i>, which collapses on both size scales. Collapsing on darkness and passivity, the pair <ʌ/u> appears affected by (non-)salience and vitality, and <ʌ/a> by difficulty (it collapses on the 'difficult' subset of the non-salient scales: passivity and smallness).

Given the widespread distribution of non-significance combined with the lack of clusters and smaller groupings of collapses, it can be argued that the three cases where a stimulus pair collapses twice requires special consideration (as a kind of mini-cluster). The <u/i> collapse does indeed repay scrutiny, in that it forms part of the unusualness of the Spanish size results. However, close attention to the <ʌ/a> and <ʌ/u> collapses appears to be less rewarding; for instance, unlike the Japanese <ʌ/i> cluster, neither appears to provide evidence for the pronunciation of <ʌ>.

10.2.1. Brilliance dimension

Following the form of the two preceding chapters, Figure 10.2 below extracts the brilliance dimension data previously presented as Figure 10.1's overall summary of the data for Spanish speakers and combines it with the frequency matrices from which it

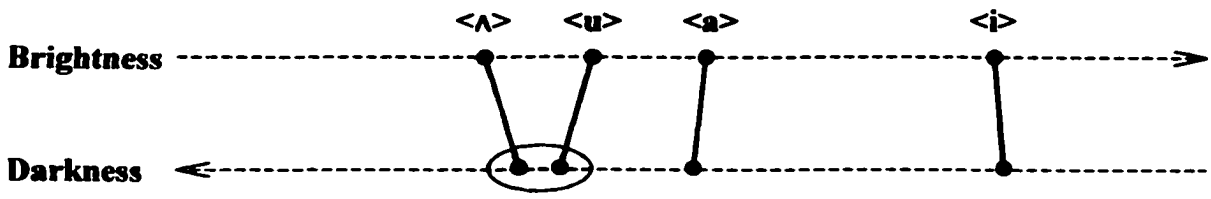
---

[100] Exceptions are AE, Korean and Spanish darkness scales (see Appendix 4). It has already been argued (see Chapter 9, Section 9.2) that brilliance appears generally to have been the easiest of the dimensions on which to make judgments, size the most difficult, and energy the most sensitive to salience.

[101] We note that while there is a widespread distribution of non-significance over stimulus pairs, the reverse is true for the scales. Passivity yields four collapses, (the maximum number, found only on one other scale: Japanese smallness) and smallness, three. In contrast, brightness and activity show none, and largeness and darkness, one.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

derives, permitting a detailed consideration of subjects' evaluations of the stimuli on the brightness and darkness scales.



Brightness

	<u>	<a>	<i>
<ʌ>	59 38	69 29	80 20
<u>		60 35	77 22
<a>			79 20

Darkness

	<u>	<a>	<i>
<ʌ>	44 52	37 62	16 84
<u>		39 59	17 82
<a>			26 72

Figure 10.2, Brilliance dimension data arrays (Spanish speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing significantly, p>.05.

General remarks on the scales. The upper panel of Figure 10.2 shows that on the salient scale rankings constitute four steps; these collapse into three on the non-salient scale, with <ʌ/u> at the darkest endpoint. Apart from this single collapse, stimuli are well spaced out on both scales, with minimal difference between their index value ranges (of .25 index points: for comparisons, see also Table 10.12). The lower panel of Figure 10.2 indicates that <ʌ/a>, <ʌ/u> and <i/a> all show somewhat stronger robustness of response on the salient scale. There is a minimal difference in favor of the salient scale for <u/a>,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

while <u/i> and <ʌ/i> show differences in favor of the non-salient scale. Thus, overall, a greater robustness of response can be claimed for the salient scale, but only a slight one.

While the Spanish brilliance rankings (including the <ʌ/u> collapse on darkness) resemble those of the AE group, locus cannot be considered alone as responsible in this case: this is because these subjects' realizations of <a> and <ʌ> are both considered to lack a locus.[102] However, the ranking of <i> as brightest/least dark can be ascribed to <i>'s high locus, together with such asssociated features as the carrying power of <i>'s high frequencies ('shrillness') and lip spreading (smile gesture), as in the AE case.[103] On the other hand, <a>'s position as second-brightest/second-least-dark is argued as due to other features: amplitude (openness/loudness), and front articulation,[104] associating it with 'bright' <i>. The ranking of <u> as third-brightest can also be ascribed to its low locus, to its backness and possibly to the damping effect of tight lip rounding[105] (i.e., its non-amplitude). The least-bright ranking for <ʌ>, rather than being due to its lack of locus (which would have associated it with <a>), can be claimed as due to its backness combined with lack of amplitude (i.e., its half-close articulation). Shared backness and lack of amplitude also accounts for the collapse of <ʌ/u> on darkness, together with the general blurring effect of non-salience.

---

[102] For arguments relating to the most likely L1 phones referred to in pronouncing the non-high stimuli, see Section 10.1.1, and for discussion of absence of locus in these phones, Section 10.1.2.

[103] For details of these arguments, and those relating to the locus of <u>, see Chapter 7, Section 7.2.1.

[104] Anterior placement for these subjects' realization of <a> is argued above, in Section 10.1.1.

[105] For the noted tightness of Spanish [u]'s lip rounding, see Navarro Tomás (1967).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

10.2.2 Energy dimension

Figure 10.3 presents the data arising from judgments on the energy scales by the Spanish subjects.

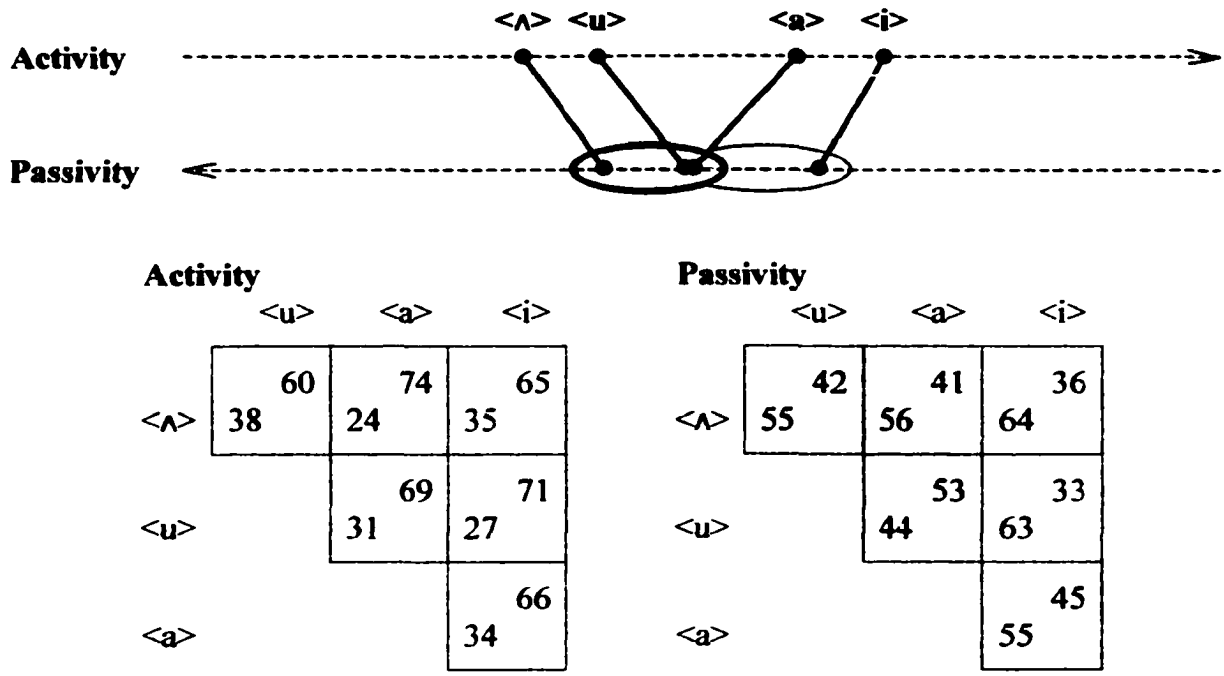


Figure 10.3, Energy dimension data arrays (Spanish speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing significantly, p>.05; bolding indicates a special case with no significant contrasts among a set of *three* stimuli.

General remarks on the scales. The upper panel of Figure 10.3 shows that on the salient scale rankings constitute four steps; these collapse into two on the non-salient scale. We also note a three-way collapse of <ʌ/u/a>on the non-salient scale, and the collapse of <a> with <i> outside this set, i.e., an indeterminate placement for <a>. There is also considerable shrinkage of the non-salient index value range (.140 index points). The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

lower panel of Figure 10.3 indicates overall greater salient-scale robustness of response; this is seen clearly in the scores of five stimulus pairs, <ʌ/a>, <ʌ/u>, <u/a>, <u/i> and <i/a>, while <ʌ/i> also shows a minimal difference in favor of the salient scale.

Energy compared to brilliance. A number of features are brought out when results for the two dimensions are contrasted. Table 10.14 below illustrates several differences in ranges.

Table 10.14: Spanish response range and discrepancy values for salient and non-salient scales of two vitality dimensions

	Spanish		
dim.	*sal.*	*Disc.*	*n-sal.*
brill.	.497	.025	.472
engy.	.350	.140	.210

Firstly, while the salient scales of both dimensions show good spacing between the stimuli (i.e., four-step rankings), we see considerable difference (of .147 index points) between their ranges. Secondly, the brilliance dimension ranges are similar for salient and non-salient scales; in contrast, energy shows a large discrepancy[106] between the two. Although rankings for all four scales are the same, passivity differs from the other three, both in its range shrinkage and in the extent of its non-differentiations: while there is a constent collapse of <ʌ/u> over both non-salient vitality scales, <a/i> and <u/a> uniquely collapse on passivity. As noted above, this suggests that brightness and darkness were

[106] See Table 10.12 above, indicating maximum response-range shrinkage for energy for all four groups.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

almost equally easy scales for these subjects to respond to, that activity was more difficult and passivity, still more difficult.

Why? In accounting for these differences, firstly we note that the applications of the two sets of vitality adjectives are different in scope. *Bright* and *dark* are conveniently wide, since they can literally describe inanimate objects (a dark room, a bright shirt) and also, metaphorically, animating/animated phenomena (a dark mood, a bright glance). In contrast, *active* and *passive* are more restricted, being applied either literally to animate entities (a passive group of students, an active baby) or metaphorically, to animating/ed phenomena proceding from them (a passive expression, active feelings). Thus, they may have seemed especially inappropriate modifiers for syllables, a type of inanimate object, (and even more so for these subjects, who were intellectually unsophisticated and lacked confidence in their English skills).[107] Secondly, all three salient terms contrast presuppositionally with the non-salient terms.[108] Thus, questions of the type *Of stimuli x and y, which is the more passive?* may have been the most difficult to answer on two counts: the adjective violated subjects' expectations about what it should be applied to, and it presupposed that either or both stimuli should be passive, which may have violated subjects' intuitions.

---

[107] For details of these subjects' relative inexperience with English compared to the other ESL groups, see Appendix 2.

[108] Thus, *How bright/active is x?* is a neutral enquiry, i.e., does not presuppose any degree of brightness/activity. In contrast, *How dark/passive is x?* is non-neutral and does presuppose some degree of darkness/passivity. In the case of a stimulus or stimuli which are judged as bright or active, questions using the non-salient terms would therefore involve violation of presuppositions and could prove especially confusing (Dr. Dianne Bradley, personal communication).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Interpretation of stimulus rankings. Generally, it can be argued that the rankings of the stimuli are triggered by the same features as on the brilliance dimension. Thus, <i>'s high locus and linked other characteristics (shrillness, lip spreading and high front tongue articulation) can again be held responsible for its endpoint most-active/least-passive positions, while the second-most-active ranking for <a> can be explained by its amplitude. When the two vitality subdimensions are compared, <i>'s more extreme position on the brilliance scales indicates that it is considered much brighter/non-dark than it is active/non-passive, and this goes well with an explanation appealing to its high frequency and shrillness etc.: these features can be argued to connote brightness better than they do activeness.[109] On the other hand, <a>, which falls in the mid-range on the brilliance scales, moves rightwards toward <i> on activity, and collapses with <i> on passivity. This can be explained in the opposite way, as <a>'s amplitude (loudness/wide jaw opening) better connoting activeness than brightness for these speakers.[110] When combined with the blurring effect of non-salience, the lessened effects of their respective triggers produce a feature equivalence: <a>'s amplitude is judged as equally non-passive as <i>'s high locus/shrillness.

At the other end of the scales, the rankings of <u> and <ʌ> can be explained, respectively by the low locus of <u>, its back articulation and damping (due to lip rounding), and by

---

[109] Thus AE index values for <i> are higher on brightness than activity, suggesting the same thing (see Appendix 3).

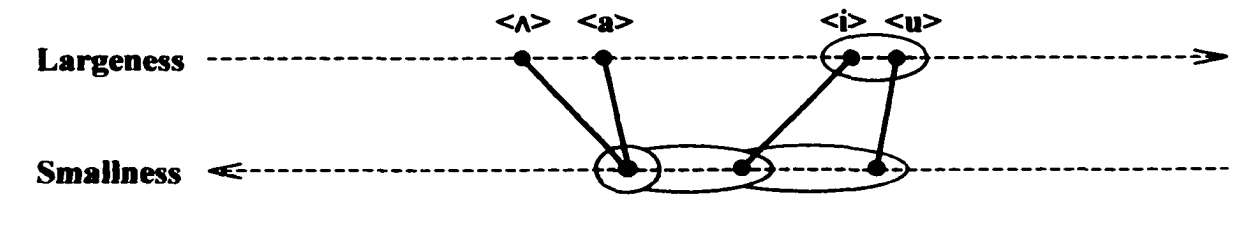
[110] We note different distributions of <a/i> collapses in two other groups' data. Found on both Korean brilliance scales, they suggest that amplitude and locus equally connote brightness/non-darkness, whereas amplitude 'wins' over locus on the energy dimension. In contrast, the <a/i> collapses on all three non-salient Japapanese scales suggest more general sensitivity to the blurring effects of non-salience.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

<ʌ>'s back, half-close articulation. The <u/ʌ/a> collapse can be accounted for as follows: <ʌ/u> share back articulation and possibly non-amplitude (<ʌ> as a half-close vowel, <u> as close and damped, i.e., lacking <i>'s shrillness); <ʌ/a> share a lack of locus. The shared features of <u/a> are more difficult to pinpoint. One possibility is that the passivity scale brought out common equivalence of articulatory spaciousness (<u>'s horizontal, <a>'s vertical), and this especially connoted passsivity. Another is that <a>, which participates in two collapses, as member of a non-passive and of a passive triad, has a feature-conflict. On the one hand, it has the 'activeness' features of amplitude and anteriority (or non-backness) and on the other hand, the 'passiveness' relation to least-active (lax, backed) <ʌ>, including a shared allophonic relation to /a/. A third possibility is that there is no coherent explanation. As argued above, applying *passive* to stimuli may have produced more confusion than phonosymbolic intuitions for Spanish subjects (we note that every stimulus participates in at least one collapse on passivity), and this confusion may have included uncertainty in comparing <a> and <u>.

10.2.3. Size dimension

Figure 10.4 presents the data arising from judgments on the size scales by the Spanish subjects.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

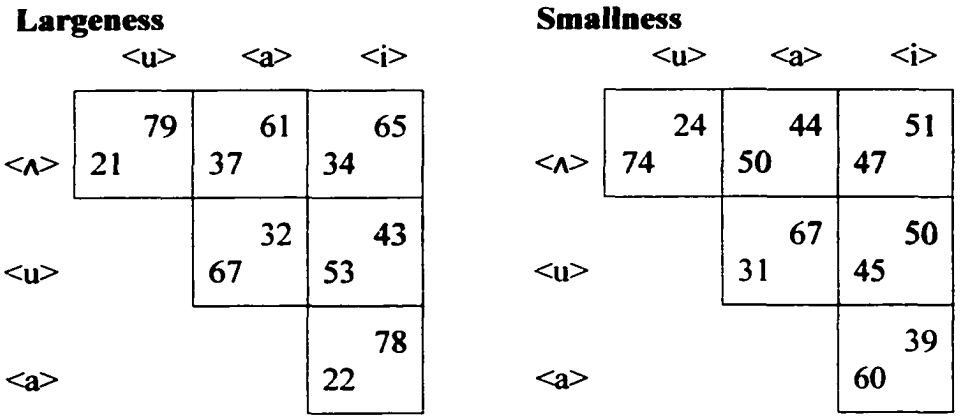


Figure 10.4: Size dimension data arrays (Spanish speakers). *Upper panel:* Ranking index values and consensus data. *Lower panel:* Elective choice frequencies ('neither' responses omitted) for all stimulus pairs. Shading indicates stimulus pairs not differing significantly, p>.05.

General remarks on the scales. The upper panel of Figure 10.4 indicates that rankings are collapsed into three steps on both scales, and in addition, that <ʌ> and <i> are indeterminately placed on smallness. The shrinkage of the non-salient index value range (.119 index points) is large.[111] The lower panel of Figure 10.4 shows overall greater salient scale robustness of repsonse; this is seen for five out of the six stimulus pairs, <ʌ/a>, <ʌ/i>, <ʌ/u> , <u/i> and <i/a>, while the scores of the sixth, <u/a> indicate equivalent robustness on both scales.

Size compared with brilliance and energy. Size judgments resemble those of energy, and contrast with those of brilliance, in being affected by salience. Table 10.15 below reproduces the Spanish portion of Table 10.12; for energy and size, we see similarities in

[111] For interlinguistic comparisons of response-range shrinkage, see Table 10.12 above.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the range values of both salient and both non-salient scales, while values for the salient and non-salient brilliance scales are similarly large.

Table 10.15: Spanish response range and discrepancy values for salient and non-salient scales of three semantic dimensions

	Spanish		
Dim.	*Sal.*	*disc.*	*n-sal.*
Brill.	.497	.025	.472
Engy.	.350	.140	.210
Size	.365	.119	.246

In addition, we have noted a resemblance between the non-salient energy and size scales in terms of their widespread collapsing, although their configurations are not the same. But there is also a configurational resemblance between the non-salient energy and salient size scales,[112] which suggests unusual ease of largeness judgments for this group. Two factors have been argued as causing these features of the Spanish data. Firstly, the relative unsophistication of the subjects may have meant that both energy terms were particularly difficult to respond to, and this therefore differentiated the clarity of responses (spacing-out of the stimuli, response-range width) on the energy scale from those of brilliance. Secondly the probable understanding of *large* as meaning *long* made salient scale size responses easier and less subject to conflicting criteria than for the other groups.

[112] See, above, Section 10.2. where it is noted that the salient size and non-salient brilliance scales have symmetrically similar configurations, i.e., share (unusually) fairly good spacing.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Interpretation of stimulus rankings. We have already remarked on the rankings of the collapsed <i/u> as largest/least small stimuli; this includes the only salient scale nondifferentiation in the Spanish data. The opposite-endpoint collapse of <ʌ/a> on smallness is striking.[113] What makes the size rankings and collapses unique in the data of all four groups are the evaluations of <i> as large/non-small and <a> as non-large/small, exactly the opposite of what was predicted and then observed by Sapir (1927, 1929).

Two factors[114] have been suggested as influencing these unusual placements. First is the the reinterpretation of the tenseness/diphthongization of AE [iy] and [uw] as *length*, thus contrasting them with non-diphthongized AE [a], which is unlikely to have been reinterpreted in this way. Second is the influence of stimulus spelling, i.e., the 'long' *hVV* spellings of the high vowel stimuli , vs. 'short' *hVh* ones for the non-high stimuli.[115] In addition, for the salient scale, a third factor is the likely interpretation of *large* as meaning its Spanish cognate 'largo,' thus licensing subjects to evaluate on the basis of perceived vowel length.[116] It has been claimed, both historically and in discussion of the findings in this experiment, that there is an iconic tie between the perceived length of a vowel and phonosymbolic perceptions of largeness. [117] Furthermore, even if subjects understood

---

[113] However, the collapse of <ʌ/a> is also found on passivity, and thus may have been partly triggered by especial difficulty of judgment argued for both scales.

[114] For details, see above Section 10.1.1.

[115] Here, it may have been a question rather of shortening the nonhigh vowel stimuli than lengthening the high vowel stimuli, due to the vowel-checking influence of final *h* (see above, Section 10.1.1).

[116] Although the smallness of the error number (21) means that this is only provable for 21% of the total subject pool, since it represents 80% of the sample, it has been argued that a similar proportion of the total pool are likely to have made the same error (see above, Section 10.2).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

large to mean both 'long' *and* 'big,' it can be argued that length (although only one subdimension of largeness, along with breadth/width and height) is the most efficient of the three in supplying a phonetic feature that will connote large size.

There are at least three articulatory features that *broad/wide* could be tied to: tongue-width (as in the difference between [s] and [š]),[118] lip-spreading, as in AE [iy][119] and jaw lowering.[120] While 'broad' is an adjective used in characterizing aspects of at least one language: English (as in 'a broad accent'), given the array of possible phonetic referents, it would appear difficult to use in a consistent way for phonetic descriptions.[121] There are two acoustic features that *high* could be connected with, i.e., two types of frequency

---

[117] For early research indicating a relation between perceived vowel duration and phonosymbolic size, see Sapir (1929: 234) and Newman (1933: 68) and Chapter 2.of this dissertation. See also discussions of size scale data in the Chapters 7 and 9, Sections 7.2.3, 82.3 and 9.2.3. For data on inherent vowel length in AE, see Peterson and Lehiste (1960) and for Japanese, see Strange et al. (1998). Korean subjects also evaluated <a> as the largest stimulus. Although I have not been able to obtain data on inherent vowel length in Korean, it is very likely that Korean [a] has similarly very long or the longest duration of the vowels, since inherent length appears to be due to wide jaw opening.

[118] This may, in terms of vowels produce a difference between the low vowels, such as [a], with the tongue flattened out covering the entire area of the lower jaw, and the high vowels such as [i] and [u] where the tongue is narrowed or buncehd to fit into the dome of the palate (Dr. Lawrence Raphael, personal communication).

[119] Yet taken to its logical conclusions, in American English this would make [iy] broad and [uw] narrow, which seems counter-intuitive. In addition lip-spreading raises the frequencies of acoustic energy on whatever vowel it affects (Tartter 1980), and in at least two data sets in the present experiment, high frequency concentration has not been found to correlate with phonosymbolic largeness, but rather the reverse. Moreover, if lip gesture did function phonosymbolically in this way, it would have to be language-specific; for example, Japanese [i] is not spread, nor is its [ɯ] rounded.

[120] This would involve the confusing translation of a horizontal dimension to a vertical phonetic feature, but might be the best candidate, particularly if tongue-width turns out to correlate with wide jaw opening.

[121] One language-specific example of the use of "broad" as a term of phonetic description can be found in an English novel by John Braine, *Room at the Top* (1957: 36). The theme is social mobility, and there are a number of references to the class connotations of features of Northern and Southern English dialects. When Braine's hero refers to "the broad *a* and the narrow *u*" i.e., the [a] of *calm* (rather than the [ae] of *cat*) and the [ʌ] of *cup* (rather than the [ʊ] of put), we understand that what is articulatorily anterior or central is 'narrow' and noncentral or nonanterior, 'broad.' But we note that this does not relate directly to any of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

information. The first is F_0 (intrinsic pitch) and the second is locus. But the height-frequency relation tends to indicate the reverse of the size of the frequency values (whether of F_0 or locus) with 'high' connoting *small* and 'low,' *large*.[122] Thus, this dimension also appears unsatisfactory for judging vowels.

In contrast, *long* can be directly (rather than inversely, as with *high*) connected with vowel duration; *duration* and *length* are synonymous in normal speech. Moreover, perception of phonetic length is richly supplied, via three sense modes: aurally (duration of vibration) proprioperceptively (duration of tongue, jaw and lip gestures), and visually (duration of jaw and lip gestures). To conclude, because of the especial aptness of length out of the three subdimension of largeness to symbolize magnitude, it can be claimed that the Spanish *large* – 'largo' mistranslations activated what was already a potential connection between the phonetic and the phonosymbolic, i.e., brought what was a subliminally-occurring relation closer to, or into, conscious awareness.[123]

Size-triggering features. In the three preceding data analysis chapters, a survey was made of the potential size-connoting features, and this approach will also be taken for the Spanish size results. However, since vowel-length perceived on a non-intrinsic basis is argued here as a major factor—possibly the dominant trigger—in size judgments, two

---

three possible phonetic correlates of breadth discussed above.

[122] Further problems with *high* would be conflicts between evaluating according to F_0 (which would equate with high tongue position, consequently giving the high vowels equal largeness) vs. according to locus (which would separate the high vowels).

[123] For discussion of the issue of subliminal phonetic-phonosymbolic connection and conscious awareness, see Fónagy (1979) and Appendix 6.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sets of values will be given for perceived duration. The first is based on intrinsic duration, and thus relates to openness/tongue height, and the second, on the explanations given above (Spanish-influenced reinterpretation of AE phonology and stimulus spellings, and the *large* – 'largo' mistranslations). As seen for the corresponding Korean and Japanese tables, there are gaps. Firstly, we note that since no acoustic data were located for the allophones of /a/, [a˂] and [ɐ] (argued to underly, respectively Spanish pronunciations of <a> and <ʌ>), all value-assignments for <ʌ> are tentative, as are those related to the anteriority of <a>. Secondly, no acoustic studies of intrinsic intensity were found for Spanish vowels; these have been assigned tentatively on the basis of openness/tongue height. Lastly, as with vowel devoicing on the equivalent Japanese table, an additional, Spanish-specific size feature is included; this is morphemic marking for <a> and <i>, respectively, for largeness and smallness. [124]

Thus, for articulatory spaciousness, <i> is assigned a value of 1, <u> (having 'long tube,' horizontal magnitude) a value of 2, <a>, 4 (based on vertical magnitude), and and <ʌ> a value of 3 or 4 depending on the importance of its non-anteriority (i.e., horizontal magnitude) vs. its half-openness. For intrinsic perceived duration, we see two sets of contrasting values. In those following the findings of del Carmen McNair (1996), the high

---

[124] Among others, Jesperson (1933) has indicated the use of the diminutive morpheme *–ito* in Spanish as one of many pieces of evidence for the generalized Indo-European association between [i] and smallness. Since *–ito* has widespread, obviously expressive use in the language (and furthermore, its first vowel [i] always comes in the tonic penult position, i.e., is metrically highlighted) a Spanish-specific attribution of smallness can therefore be argued for <i>. I am also informed (by Dr. Ricardo Otheguy, personnal communication) there are two morphemes opposed to *–ito*: *–ote* is the simple aggrandizer, while *–azo* has a slightly more specialized use, including in it the sense of "a blow" or "a strike" (in both, we note V_1 in tonic penult, metrically highlighted position). Thus, there is a less absolute morphemic association between [a] and largeness, but it also exists.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

vowel stimuli are assigned a value of 1, <ʌ>, 2 and <a>, 3, while in those following the arguments on Spanish interference given above, <i> and <u> are assigned a value of 2, while <ʌ> and <a> receive 1. Hypothesized perceived intensity values are 1 for <u> and either 1 or 2 for <i>, (depending on whether the tight lip rounding noted for Spanish [u] further reduced its acoustic intensity compared to <i>'s).[125] Parenthesized values for <a> and <ʌ> reflect this issue. Both high vowels are assigned intrinsic pitch values of 3,[126] <ʌ>, of 2 and <a>, of 1. As they are hypothesized to lack a locus, <a> and <ʌ> are assigned a value of ø, while <u> receives a value of 1, and <i> of 2. Finally, for morphemic large/small marking, <a> receives a value of 2, and <i> of 1; <ʌ> and <u> receive ø, since they are not marked in this way.

---

[125] For details of similar arguments for the damping of Japanese <u>, see Chapter 9, Section 9.2.3

[126] Del Carmen McNair (1996) reports little difference in F_0 values between [i] and [u] in either data set.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 10.16: Observed stimulus rankings for size dimension (order being specified as *least size* = 1 through *most size* = 4), compared with rankings predicted by potential size-triggering features.

Observed rankings	**1**	**2**	**3**	**4**
Largeness	**<ʌ>**	**<a>**	**<i>**	**<u>**
Smallness	**<ʌ>**	**<a>**	**<i>**	**<u>**
Feature-based rankings	**<ʌ>**	**<a>**	**<i>**	**<u>**
Artic. Spaciousness	3 (or4)	4	1	2
Percd. Duration (del C. M.)	2	3	1	1
Percd. Duration (Sp. Intf.)	1	1	2	2
Perceived intensity	3 (or 4)	4 (or 3)	2 (or1)	1
Intrinsic pitch	2	3	1	1
Energy concentration locus	Ø	Ø	1	2
Morphemic marking	Ø	2	1	Ø

Comparison between observed and feature-based rankings in Table 10.16 indicate that the <i/u> collapse and rankings on both scales cannot be ascribed to articulatory spaciousness, or to intrinsic perceived duration as based on the del Carmen McNair findings; instead they appear a product of Spanish interference (phonetic and orthographic). On the largeness scale, these effects can be argued as intensified due to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

lexical (cognate) interference, which accounts for the unique salient scale collapse (rather than the independent ranking of <u> as larger than <i>, due to its greater horizontal articulatory spaciousness.Normally, <i/u>'s most obvious common articulatory characteristic of tongue height/jaw closeness, together with its acoustic correlates (shorter duration and higher F_0),[127] would lead to expected judgments of smallness. Indeed, the association between articulatory nonspaciousness and phonosymbolic smallness (modified in the case of <u> by its backing/rounding) has been argued at various points in the analysis of the other data sets.

The ranking of <ʌ> as least-large/smallest can be similarly ascribed to lexical and orthographic Spanish interference, with effects again intensified on the salient scale due to the *largo*-'large' influence. The fact that three of the stimuli appear to be ranked according to interference-based length judgments argues that this, rather than articulatory spaciousness, was the dominant trigger for Spanish speakers, and the collapse of <ʌ/a>, and the consequent endpoint ranking of <a> on the smallness scale can also be ascribed to the same causes.

Feature conflicts and indeterminacies. On the other hand, the mid-range placement of <a> on the salient scale (as second-least-large, rather than non-large) can be claimed to reflect a conflict between its features, i.e., that the dominance of length as a criterion is not total for this stimulus. First is the fact that Spanish [a] has the *–azo* feature, i.e., is

[127] This list should also include lower intensity, though no acoustic studies dealing with this aspect of Spanish vowels could be located.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

morphemically marked for largeness. Second, it can be argued that conflicting with <a>'s interference-based shortness are those features claimed to have produced largest/least-small rankings for it in the data of the other three language groups: vertical articulatory spaciousness, loudness, low F_0 (and possibly even articulatorily-based intrinsic duration, which would have made it the longest stimulus).

A similar feature conflict for <i> can be argued as the cause of its collapse with <ʌ> on smallness, and of its consequent indeterminacy. While <i> shares (interference-based) length with <u>, it has the *–ito* morphemic smallness marking. This can be seen as pulling it toward smallest <ʌ> (a feature equivalence, respectively of shortness and marking), yet ensuring that it stays differentiated from second-smallest <a> (which is oppositely-marked from <i> morphemically, but with which, as a mid-ranked stimulus, it should otherwise also have collapsed).[128] Thus, the indeterminacy of <ʌ> can be accounted for as follows: it shares interference-based shortness with <a> and feature-equivalent smallness with <i>.

10.2.4 Concluding remarks on the Spanish data and a phonosymbolic profile of the stimulus vowels according to Spanish-speaking subjects.

Salient vs. non-salient robustness of response. The scatterplot (Figure 10.5 below), whichcorresponds to Figures 7.6, 8.5 and 9.6, respectively in Chapters 7, 8 and 9, allows

[128] A less satisfactory smallness feature for <i> is its lack of both horizontal and vertical spaciousness (see Table 10.16). But while spaciousness will differentiate <i> horizontally from <u> and vertically from <a>, it should have also differentiated it horizontally from <ʌ>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

us to see whether the greater robustness found in the salient scale responses of all three language groups also characterizes the Spanish data.

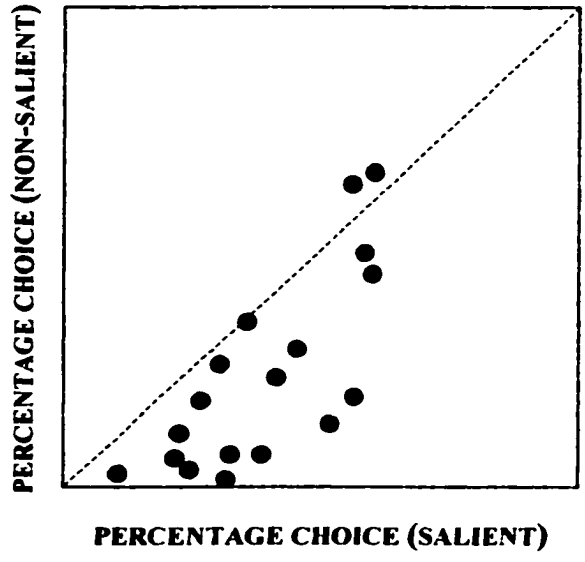


Figure 10.5: Preference strength in antonymic scales (Spanish speakers). Percentage of elective choices ('neither' responses omitted) favoring preferred stimuli for each of 6 stimulus pairs evaluated on salient vs. non-salient sub-scales of 3 dimensions. Axis values run 50 to 100%.

We note that 14/18 points are below the diagonal (indicating numerically stronger salient scale preferences for a stimulus); located on the diagonal we find 2/18 points (indicating that preferences are of equal strength on salient and non-salient scales) and above the diagonal there are 2/18 points (indicating stronger non-salient scale preferences). Not only do we find a large majority of salient scale preferences, but 8/14 salient points are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

located well below the diagonal while 2/2 non-salient points are close above it. Thus we see small majorities for the preferred stimuli on the 'winning' non-salient scales, and either similar or larger majorities for those on the 'winning' salient scales. It can therefore be claimed that the tendency found in the AE, Korean and Japanese data for a more robust response pattern on the salient scale is repeated once again in the Spanish findings, and this interlinguistic agreement strongly suggests a general trend.

Triggering features. As with the other three data sets examined, the Spanish rankings are the same for brilliance as for energy. Therefore, we once again ask, firstly, whether brilliance and energy were synonymous for the Spanish speakers and, secondly, whether the same triggering features were operating on both dimensions. A highly qualified yes answers the first question. Although rankings were the same, findings on the dimensions diverged in terms of their responses to salience; this is seen in differences between the configurations on the non-salient scales and in the different values of the discrepancies between salient and non-salient scales on the two dimensions. In addition, there was a large difference in response ranges values for both salient scales, suggesting that as an evaluating measure energy was considerably more opaque in application than brilliance, and that salience added a further level of opacity to energy judgments. It appears that the same triggering features—locus and amplitude—were operating in judgments on both dimensions, but that locus connoted energy less well than it did brilliance, while amplitude connoted it better.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The main triggering feature for size is argued to have been perceived duration, based on various types of Spanish interference not forseen in the experimental design. Although this produced incidental evidence suggesting duration can be a highly efficient connotor of size, the issue here is principally one of experimental foresight. In future studies of this type (i.e., ones involving speakers of languages related to the investigatory language), word lists and scale terms should be screened for cognates. Moreover, regardless of the subjects' first language(s) if stimuli are spelled, unintended features (such as potential short-long oppositions) have to be checked for and eliminated. Another size trigger, conflicting with duration, is argued to have been the Spanish morphemic marking of two of the stimulus vowels; again, this is an aspect of stimulus/scale choice which must be controlled for. Clearly, if a stimulus vowel is also found in a morpheme denoting an attribute judged on a scale, this may 'color' the vowel and skew findings on that scale.

Phonosymbolic profiles of the stimuli.

<a> Non-consensus: 4/18 (22%). This stimulus is mid-ranked on the vitality scales and largeness and at an endpoint on smallness. Its differentiability is affected by scale (it collapses with all three other stimuli on passivity). It is also affected by pair-partner and difficulty (it collapses with <ʌ> on passivity and smallness, the 'difficult' subset of the non-salient scales). It can be described as neutral with regards to brilliance, somewhat active and small.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

<i> Non-consensus: 4/18 (22%). This stimulus is ranked at an endpoint on the vitality scales (independently on all but passivity) and mid-ranked on size, toward the large/non-small end of the scale. Its differentiability is affected by dimension and pair-partner (collapsing with <u> on size), and by salience/difficulty (it collapses with <ʌ> on smallness and <a> on passivity). It can be characterized as vital, and ambiguous or conflicted with regards to size, though with stronger tendencies to largeness than smallness.

<u> Non-consensus: 5/17 (28%). This stimulus is ranked at an endpoint on the size scales and is mid-ranked on the vitality scales, toward the non-vital ends. Its differentiability is affected by dimension and pair-partner (collapsing with <i> on size), by vitality/salience and pair-partner (collapsing with <ʌ> on darkness and passivity), and by scale (it also collapses with <a> on passivity). It can be described as non-vital (somewhat more dark than it is passive), and large.

<ʌ> Non-consensus: 5/17 (28%). This stimulus is ranked at an endpoint on all 6/6 scales. Its differentiability is affected by salience, since its placements are independent only on the salient scales. More specifically, it is affected by vitality/salience and pair-partner (collapsing with <u> on darkness and passivity) and salience/difficulty and pair-partner (collapsing with <a> on passivity and smallness). It can be described as non-vital and small.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Since there are minimal differences in the number of non-consensus responses for the four stimuli, they have not been characterized in terms of how well-defined/separate their phonosymbolic 'personalities' are. However, it is possible that <a> and <i> could be described as somewhat better-defined than <u> and <ʌ>. Minimal variations in stimulus differentiability in the Spanish data might be interpreted as corresponding to greater differentiability variation in the other three data sets.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 11

PHONETIC SYMBOLISM ACROSS FOUR LANGUAGES

We now move from interpreting each language's data in isolation to an across-the-board consideration in search of commonalities and divergences. The first issue to be treated is whether findings provide statistically significant evidence for the existence of phonetic symbolism. Section 11.1 presents evidence which does indeed support its existence; it then looks beyond this to consideration of whether minor variabilites between languages has any significance. Section 11.2 brings together two kinds of data and discusses these in terms of the contributions made by language-specific and by universal phonosymbolisms, together with the part which may have been played by experimental design in instances of either type of response. In order to explain response patternings, a number of phonosymbolic triggering features are proposed; this section concludes with a comparison between these and the triggering features proposed by Sapir (1929) and Newman (1933). Finally, section 11.3 treats the question of whether any cross-linguistic generalizations can be made about the phonosymbolic 'personalities' of each of the four stimulus vowels.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

11.1 The existence of phonetic symbolism

An initial analysis, at the broadest possible grain, asks whether the extent to which responses to the items of the questionnaire showed consensus[1] provides a statistical argument that there was indeed a systematic basis for subjects' behavior. Table 11.1 summarizes data in these terms, presenting for each language group the total number and percentage of the 36 questionnaire items which, by individual test, showed uniformity in the responses at a level greater than would be expected by chance. In order that the case to be made is appropriately conservative, instances of trend are *not* included in the totals, but for completeness are cited in the table (italicized and in parentheses).

Table 11.1: Total numbers of items evoking consensus (maximum 36)

Speaker Group	Number of Questions	Percentage
English	29 *(1)*	80.5%
Korean	25 *(3)*	69.4%
Japanese	23 *(3)*	63.8%
Spanish	27 *(0)*	75.5%

For any individual statistical test, the machinery of hypothesis testing fixes the probability with which an outcome will be claimed as 'significant' when in fact the null hypothesis is true. This is the level of significance of the test, set here (as per convention) at $\alpha = .05$. As is well-known, the probability of such 'false alarms' increases

[1] The reader is reminded here that the term 'consensus' refers to a response pattern in which subjects within a language group agree sufficiently often (about which of two stimuli had more of a named attribute) that the outcome differs, $p<.05$, from the 50/50 split predicted by chance. See Chapter 6, Section X.X for details.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

rapidly when we consider a *set* of individual tests, taken together; and for the case of a set of 36 tests, in particular, the probability of at least one false alarm is very high ($p = .84$). The statistical argument that subjects' responses have some systematic basis must therefore rest on the fact that not merely a scattered few but rather a *substantial majority* of questions were responded to with consensus. In Table 11.1, we see that number of questions for which consensus was exhibited ranges from 23 (Japanese) to 29 (English), for an average of 26 out of 36 questions.

If we consider the set of questionnaire items taken together, for any language group, the likeli*hood* that a given number of false alarms occurs is specified by an asymmetric binomial distribution with parameters $p = .05$, $q = .95$, $N = 36$. The properties of that distribution provide firm assurance that the numbers of consensus responses observed is not a matter of chance, but rather solid evidence of the existence of phonetic symbolism. To illustrate: 6 or more consensus instances would purely by chance be identified in a set of 36 tests with probability $p < .01$, 8 or more, with probability $p < .001$, and 9 or more, with probability $p < .0001$. Clearly, the likeli*hood* that chance alone would lead to the numbers of consensus outcomes summarized in Table 11.1 is vanishingly small, orders of magnitude beyond the levels conventionally set for statistical significance. Thus for each group, the data can surely be said to indicate the presence of a consistent phonosymbolic response.

A subsidiary question concerns the language-to-language variation in the numbers of questions answered with consensus. Is that variation itself significant? A chi-square test of independence shows that it is not, $X^2(3) = 0.77$, $p > .50$. This lack of significant

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

variation in the extent of consensus among the four groups is what we should anticipate. If a phonosymbolic response exists, there is no reason to expect it to be confined to American English or any other group of speakers. However, no claim is being made here that phonosymbolic evaluations were being made on the same basis from group to group. Instead, the assertion is only that (at least at this grain of analysis), for whatever bases the language groups are using to drive phonosymbolic responses, these do not differ in effectiveness.

11.2. Universality versus language-specificity in phonetic symbolism

Although evidence for phonetic symbolism has been established for all four groups, the question is still to be answered to what extent evaluations of stimuli were agreed upon between the groups, and therefore if the findings suggest a universal or language-specific basis to the responses. For greater clarity, while the term *consensus* is used for intralinguistic accords between speakers (see Fn. 1 above), the term *agreement* will be used only for interlinguistic accords; e.g., there is agreement between English, Korean and Spanish speakers in ranking [ʌ] as the least bright and least active stimulus.

The two principal types of data, the distribution of consensus and the stimulus rankings, provide a number of partial answers to the issue of language specificity versus universality. In the rankings, we will find some notable cases of agreement, indicating universality, but there are also certain disagreements, i.e. apparent language-specific areas of phonosymbolism. Amongst these, some appear to have been produced by the reactions between a specific group and way in which the experimental design was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

implemented, while others do not seem to have been the result of experimental conditioning but rather of a free, language-specific response. Finally, there is one disagreement between groups which appears to have a cultural basis, rather than being classifiable as either language-specific or universal. The consensus data will be treated first, followed by the ranking data. This moves our consideration through a number of specific cases to more general issues of interlinguistic agreement and disagreement.

11.2.1. Consensus distribution data and universality versus language-specificity

Data from the set of four frequency matrices has combined into a single non-consensus summary table (table 11.2). showing totals for each of the stimulus pairs as evaluated by all four language groups: notable here is that breakdowns by scale indicate twice as many cases of non-consensus on the non-salient as on the salient scales. The table is constructed to highlight instances in which speakers within language groups fail to agree in the responses to a given stimulus pair for a given semantic scale. The focus is thus on where statistical testing has failed to show a response pattern differing from the 50/50 split predicted by the null hypothesis. The data are grouped in table 11.2 to show how many and which of the twentyfour comparisons (six semantic scales, by four language groups) showed non-consensus; for convenience the number of consensus instances is also listed. Non-consensus data include both instances of nonsignificance and of trend, i.e., all instances in which the conventional 5% level of significance was not achieved.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 11.2: Non-Consensus Distribution, by Stimulus Pair

PAIR	NON-CONSENSUS CASES			CONSENSUS
ʌ / u	12			12
		EN, 3	—	dark
			large	small
		KO, 5	bright	dark
			active	passive
			large	—
		JA, 2	—	passive
			—	small
		SP, 2	—	dark
			—	passive
ʌ / a	3			21
		KO, 1	—	*small*
		SP, 2	—	passive
			—	small
ʌ / i	7			17
		JA, 6	*Bright*	dark
			Active	passive
			Large	small
		SP, 1	—	passive
u / a	7			17
		EN, 4	*Active*	passive
			Large	small
		KO, 2	*Large*	small
		SP, 1	—	passive
u / i	5			19
		KO, 1	—	passive
		JA, 2	large	*small*
		SP, 2	large	small
a / i	6			18
		KO, 2	Bright	*dark*
		JA, 3	—	*dark*
			—	small
			—	passive
		SP, 1		

Note: Italics indicate cases of trend ($.05 < p < .10$) to consensus

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 11.2 indicates that consensus on the differentiability between stimulus pairs was generally good; in most cases, for each language group a stimulus pair failed to evoke consensus no more than once or twice. A small amount of non-consensus was expected (one pair on one scale per group), but anything above that amount was considered potentially an unusual case, and thus as possible evidence of either a language-specific or universal phonosymbolic response. The most striking cases were *clusters* of non-consensus (one language group's failures of consensus on a stimulus pair over five or six scales). Then came *knots* of non-consensus (one group's failures of consensus on a stimulus pair over three or four scales). Finally were cases which appeared to be on the cusp between normal and abnormal consensus distribution; these were *small knots* of non-consensus (consensus failures on a stimulus pair over two scales by a language group). We should note here that the Spanish group has at most small knots, suggesting that, for at least one set of subjects, small knots constituted marked instances of non-consensus. In all, a total of twelve cases were found; interpretation of the consensus data means explaining these twelve cases.

Table 11.3: Instances of non-consensus requiring interpretation

	Clusters	Knots	Small Knots
English	—	[ʌ/u], [u/a]	—
Korean	[ʌ/u]	—	[u/a], [a/i]
Japanese	[ʌ/i]	[a/i]	[ʌ/u], [u/i]
Spanish	—	—	[ʌ/u], [ʌ/a], [u/i]

In each case, the question is whether the knot or cluster can be isolated as the reaction of a single language group, suggesting language-specificity of response, or whether it spans

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

language groups, suggesting a more universal failure to differentiate. But in addition, some instances of non-consensus may be due to difficulties caused by experimental design rather than to phonosymbolism. For example, six out of the twelve cases (one of the Japanese and two of the Spanish small knots, one of the English knots and both of the clusters) involve the central stimulus vowel [ʌ]. How many of these may be ascribed to the potentially problematic inclusion of *HUH* as a stimulus (confusing spelling, lack of first-language phoneme equivalent for ESL subjects, and so on)?

Non-consensus on [ʌ/u]: A case of negative agreement. We see that in table 11.2 , for five out of the six vowel pairs, differentiation is good (between 17/24 tests and 21/24 tests show statistically significant consensus), but there is one large area where differentiation is poor; it covers the overall responses for the pair [ʌ/u] (where 12/24 tests fail). Moreover, a non-consensus cluster can be found within the total of non-consensus responses for this stimulus pair: Korean speakers evaluated [ʌ/u] without differentiating the pair on 5/6 scales, a unique pattern of response in this group's consensus data overall. The reason for this may be either an allophonic relation between the vowels of [ʌ] and [u] because of Korean pronunciation difficulties with [ʌ] (i.e., a language-specific response, but the result of a problem in the experimental design) or the dominance of the sense of a shared feature (backness) common to both vowels as pronounced by Korean speakers (i.e., a true language-specific phonosymbolic response).

It could be claimed that the Korean data are skewing the overall total for this pair, and that the Korean responses constitute a language-specific case of poor differentiation. However, even without the Korean cluster, the lack of differentiation for [ʌ/u] is still

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

strikingly widespread; it would seem that all four groups had some difficulty in differentiating between the members of this pair. We note first that English speakers failed to differentiate [ʌ/u] on 3/6 scales. For this group, there seems to be no doubt that [ʌ] was a separate phoneme, or that it should have been clearly indicated via the stimulus spelling *HUH*, since ***huh*** is the conventional way of orthographically representing this syllable in English. In addition, [ʌ/u] evoked a small knot of non-consensus from both Japanese and Spanish speakers. But the most important fact is that non-consensus on this pair is so widespread: each of the cases takes on an added emphasis due to the presence of all the others. Thus, although the Korean cluster may have been the product of a mispronunciation of AE [ʌ] (and therefore an artefact of the experimental design), it is embedded in what appears to be a much larger case of genuine interlinguistic agreement.

The English knot of non-consensus on [ʌ/u] is a strong piece of evidence that the differentiation problems for this pair were not conditioned by being an ESL speaker, but instead *by a genuine lack of contrast between the stimulus vowels when paired.* The question is now whether those aspects provoking the widespread non-consensus are the product of real interlinguistic phonetic symbolism or of some problem in experimental design. There are two different answers, indicating both possibilities. One explanation for the agreement is that there were genuine cross-linguistic intuitions of a lack of difference between the vowels, based on commonality of subjects' pronunciations. The other is that there was an orthographically based confusion between the stimulus vowels.

The *U* in the stimulus spelling *HUH* may have been misread as indicating the high back vowel [u] by some of the subjects, since AE [ʌ] and [uw] are often designated by the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

same letter (as in *tub* and *tube*). A visual-phonetic connection could thus be made between both vowels by the inexperienced reader of English, activated by those questions concerning the stimulus pair [ʌ/u]. Such a misreading of *HUH* is more likely for the non-native speakers as less experienced readers of English, i.e., as more liable to build up word pronunciations by phoneme-letter associations rather than by whole-word recognition. Thus, some part of the [ʌ/u] non-consensus may have been the result of a flaw in the experimental design (confusing stimulus spellings) and not of a phonosymbolic response which spanned all language groups.

But it can also be argued that a large part of the interlinguistic non-consensus (even excluding the Korean cluster as an ambiguous case) may be a genuine instance of agreement, a real universal—but a negative one, especially apparent on the non-salient scales.[2] *All the subjects are found to agree on the difficulty of distinguishing phonosymbolically between these stimuli.* Such an explanation suggests that *huh* was still pronounced by all ESL subjects with a fair degree of closeness to English [ʌ], despite the impreciseness of the pre-questionnaire pronunciation word-list procedures for this vowel. Whatever Korean, Japanese or Spanish phonemes the stimulus [ʌ] was associated with, its phonetic realization may well have shared a kind of central articulatory 'grey area' with AE [ʌ]. Many or all of the [ʌ]-hybrids produced by the ESL speakers *and* the timbre of [ʌ] as produced by the English speakers could therefore have had a common centrality of articulation, and a consequent acoustic "muddiness" (Masson, 1952). Thus, in the right context (i.e., the brilliance and energy scales) the muddiness of [ʌ] and [u]'s attributed

[2] The non-salient scales are those where the most non-consensus is found, i.e., the areas of greatest phonosymbolic indistinctness.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

qualities of darkness and passivity may have become phonosymbolically associated or confused with each other by subjects, whether or not they spoke English as a first language.

In the end, it can be said that both the orthographic element and the phonosymbolic play their parts in the widespread non-consensus over [ʌ/u], but that there is no way to separate them or decide which element prevails. The only honest conclusion is that some unquantifiable part of the results for [ʌ/u] is due to negative interlinguistic agreement. To find out how much, the stimuli would need to be retested with more efficient spellings.

Other cases of non-consensus involving [ʌ]. The above explanations cover four out of the six cases of non-consensus involving [ʌ], but two more remain: the Japanese cluster on [ʌ/i] and the Spanish small knot on [ʌ/a]. I shall argue that both of these are language-specific, but due to experimentally-engendered difficulties, and not to phonetic symbolism.

The Japanese cluster has already been analyzed in Chapter 9, Section 9.1.4. To recap, it was explained as the result of splitting the subjects' first-language phoneme [ɯ] into two allophones, the fronted and unrounded [ʉ] and the backed rounded [u], in order to find first-language equivalents for the vowels of the stimuli [ʌ] and [u]. However, in the polarization of these two allophones, the fronted unrounded variant [ʉ] approached dangerously close to the Japanese high front vowel [i] (which is not accompanied by lip-spreading as it is in English). Thus, neither anteriority nor lip gesture distinguished the Japanese version of the two stimuli [ʌ] and [i]; phonetic confusion led to phonosymbolic

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

confusion, resulting in the non-consensus cluster. This is definitely a language-specific case, produced by the conflict between the Japanese vowel system and the demand for an 'extra' vowel. But it cannot be claimed as an instance of language-specific phonetic symbolism.

If the proposed strategy taken by the Spanish speakers for pronouncing [ʌ] was, as described in Chapter 10 (Section 10.1.2.3), to assign it to an allophone of their [a], then it would be reasonable to expect that the pair [ʌ/a] should have evoked the same type of non-consensus clustering in the Spanish data as is found for [ʌ/i] in the Japanese data. Yet in fact, Spanish speakers failed to differentiate [ʌ/a] on no more than 2/6 of the scales: the non-salient ones of passivity and smallness (the two scales which across all groups evoked the lowest amount of consensus). This seems to argue that Spanish speakers experienced no especial difficulty in differentiating between these two stimuli, but is this truly the case?

Although the Spanish consensus-by-pair table (10.x) shows an apparently unremarkable degree of non-consensus for [ʌ/a], in contrast, the *interlinguistic* non-consensus table (11.2) shows outstandingly good differentiation overall on [ʌ/a], with consensus reached on the highest number of scales, 21/24, for this pair.[3] In this context, the Spanish results for [ʌ/a] can be seen as unusual; they suggest that the Spanish speakers were experiencing markedly more difficulty than the other groups in distinguishing between

[3] For two groups, the Japanese and the English speakers, this pair evoked consensus on all six scales, and consensus on 5/6 scales for the Koreans.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

these two stimuli. Thus the vowels of [ʌ] and [a] might well have been related to each other allophonically in the pronunciations of this language group.

Conclusions about [ʌ]. It has been observed that [ʌ]'s vowel lacks an equivalent in at least two of the three ESL subject groups' phonemic repertoires; moreover, the word list designed to clarify its pronunciation may have had the opposite effect, as could its presentation in the spelling *HUH*. Thus, it is likely to have been the most problematic of the stimuli. We see that out of the six cases of non-consensus involving [ʌ], four indicate a (negative) interlinguistic agreement, and at least some part of this agreement might be a real case of phonetic symbolism and not just a product of problematic stimulus spelling. Yet in contrast, there are two more cases which are best explained as due to the inadequacy of the subjects' first-language vowel systems to cope with the inclusion of the vowel of [ʌ]; that is to say, they are language specific, but caused by experimental design.

Where does this conflict leave us? How does the negative agreement on [ʌ/u] bear on the potentially problematic nature of [ʌ]? One explanation is that [ʌ] could have been categorized as belonging to a different phoneme by each of the four language groups, while its pronunciation remained reasonably stable and similar from one group to another. In this case, both the phonemic alliances *and* the phonetic qualities of [ʌ]'s vowel would have had their phonosymbolic effects, depending on the context (scale, language group, stimulus paired with it). The fact of two vowels' being allophonically related was responsible for the Japanese non-consensus on [ʌ/i] and the Spanish non-consensus on [ʌ/a], and may possibly have caused the Korean non-consensus on [ʌ/u].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

However, in this last case, the agreement on inability to differentiate between these stimuli was so widespread that the Korean cluster could well have originated in a phonetic, and therefore genuinely phonosymbolic, confusion between the two stimulus vowels.

Remaining cases of non-consensus. The six remaining cases of non-consensus are the English knot and Korean small knot on [u/a], the Japanese and Spanish small knots on [u/i], and the Japanese knot and Korean small knot on [a/i]. Contrasting with the general good differentiation on [u/a], the size of the English knot (4/6) suggests that the English response is indeed language-specific. Thus, it gives one piece of evidence against the view that where there is interlinguistic agreement, it is because the ESL speakers' responses were contaminated by their knowledge of English and/or by the use of English in the experiment.[4] What argues strongly in favor of language-specificity here is the distribution of English speakers' non-consensus for [u/a]: it is found on both energy and size, two dimensions evoking dissimilar rankings. This distribution suggests phonosymbolic confusion between these stimuli much more strikingly than if all four cases of non-consensus had been found on the dimensions of energy and brilliance, which *do* evoke similar rankings in the English data (as indeed they do across all groups).

However, as observed, there is also a small knot of non-consensus on [u/a] in the Korean data, both cases occurring on the size scales. Therefore, although there is wider non-consensus for this stimulus pair in the English data, it can be claimed that there is *some*

[4] Given that the questionnaire and administration procedures were all in English, and that the ESL subjects were almost invariably students of English, it might be claimed that the knowledge of English was acting as a unifying influence on phonosymbolic responses.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interlinguistic agreement here; moreover, it is between speakers of unrelated languages. It suggests that attempting to contrast [u] with [a] could prove generally problematic in size judgments.

The Spanish small knot on [u/i] (on both the size scales) indicates language-specific phonetic symbolism, but provoked partly by lack of foresight in the experimental design and partly by Spanish-to-English phonetic interference. It is the result of three factors: a) an all-too-easily mistranslatable size epithet, b) misleading stimulus spellings, and c) a phonetic misinterpretation of the quality of the diphthongized English high vowels by Spanish speakers.

Firstly, findings indicate[5] that these subjects mistranslated "large" as *largo*, that is "long," due to the presence of the cognate form in the first language. Even though "small" was correctly translated as *pequeño*, subjects did not check the asymmetry of terms and retranslate, probably because of their lack of sophistication with English.[6] Secondly, it is likely that Spanish subjects (whose native vowels are all nondiphthongized and of phonologically equivalent length) phonetically reinterpreted the diphthongized English high vowels in the stimuli [i] and [u] as longer than nondiphthongized nonhigh

---

[5] Tests on potential mistranslation were applied late in the experiment for the Spanish subjects, so these findings must be considered as suggestive rather than definitive; only 26/100 valid responses were gathered (a 27th was disqualified). Out of those 26, 21 (that is, 81%) gave *largo* as a translation of "large."

[6] 51% of the Spanish subjects reported between under one year to two years' experience with English; only 35% had five years or above experience with English. The awareness of cognate mistranslations is a fairly sophisticated and late-acquired skill for Spanish-speaking CUNY students, who made up the population of Spanish-speaking subjects in this experiment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

stimulus vowels [a] and [ʌ]. Thus, [i] and [u] become the long/large stimuli, and [a] and [ʌ] the short/small ones.

Thirdly, the spellings of the high vowel stimuli *HEE* and *HOO* featured double vowel letters (suggesting free vowels), while the nonhigh vowel stimuli *HAH* and *HUH* were spelt with single vowel letters followed by *H* (suggesting checked vowels). Authorities on Spanish phonetics agree that the vowels in free syllables are articulated with greater duration than those in checked syllables (Colton 1909, Navarro Tomas 1967 etc.). Therefore, a misinterpretation of the stimulus spellings would have added to the tendency to pronounce the high vowel stimuli with longer vowels than those of the nonhigh stimuli.

It can be concluded that all of the likely causes of the small knot on [u/i] (and indeed all of the Spanish size evaluations) are language-specific and phonosymbolic. However, it cannot be claimed that the evaluations (including the small knot of non-consensus treated here) are due to unconditioned Spanish-specific phonetic symbolism. Had the experiment been designed taking into account the problem of cognates, of phonetic interference, and of misinterpretable stimulus spellings, phonosymbolic largeness might not have been confused with length for these speakers, and their size evaluations might well have been different.

Japanese non-consensus on [u/i] also indicates a language-specific response. It is found nowhere else except size, and strikingly, is more pronounced on the salient scale of that dimension. A claim could be made that this small knot was produced by a misreading of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the stimulus spellings *HOO* and *HEE* as indicating two-morae vowel pronunciations and thus as having equivalent length—i.e., size—in contrast with the one-mora pronunciation suggested by the spellings *HAH* and *HUH*. However, this group's ranking data do not support this interpretation, and neither do their consensus data for [ʌ/a] on the size scales—or any other scales.[7] For the Japanese speakers, it seems that on the (overall better-differentiated) salient size scale both [u/i] and [ʌ/i] were opposed *as pairs* to [a], isolated as the largest stimulus: interlinguistically, a unique configuration on this scale. Thus, not only does the Japanese [u/i] size collapse seem part of this group's language-specific response to the size measures, but in contrast to the Spanish case, does not appear to be provoked by experimental design.

The Japanese knot on [a/i] is found on all three non-salient scales. In comparing the two stimulus vowels, in contrast with the other five pairs, we find neither phonetically nor phonemically are there any obvious reasons for the collapse. Instead they may be said to offer a large number of oppositions (high versus low and front versus nonfront tongue positions; main concentration of acoustic energy in the high frequencies versus main concentration in the middle frequencies; nondevoiceability versus devoiceability,[8] lack of diphthongization versus diphthongization[9]). The Japanese knot of non-consensus here may indicate a general tendency for these stimuli to collapse on those scales evoking greater non-consensus, or it may be the product of two different responses, one on the

---

[7] There are no cases of nonconsensus on this pair in the Japanese data.

[8] See Chapter 9, Section 9.1.2, for discussion of the tendency of Japanese high vowels to devoice in the environment C__C where C is voiceless.

[9] The tendency for English vowels to diphthongize would be noted only by those Japanese speakers sensitive to such a detail of English pronunciation and thus likely to be influenced by it.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

size scale and a different one on the brilliance/energy scales (which rank stimuli in the same order). Japanese smallness evoked the maximum number of non-consensus responses for any one scale (4 collapses),[10] as did smallness overall (11/24). Thus, part of the explanation may be that there was a particular tendency for Japanese speakers not to differentiate between any potentially collapsible pair on this scale.

In contrast, the Korean non-consensus on [a/i] is confined to the brilliance scales, and, strikingly (like the Japanese [u/i] case above) is more pronounced on the salient scale. Thus, there is strong indication of a link between the dimension and the collapse, which certainly suggests language-specificity. However, connections can also be made with the Japanese [a/i] knot, since rankings on brilliance and energy pattern together across all the groups, and the Japanese [a/i] non-consensus is found on one brilliance and one energy scale. If the four speaker groups are arranged according to geographic language origins, we see that the Asian language speakers (Korean and Japanese) collapse [a/i] on 5/12 scales, while the European language speakers (English and Spanish) collapse them on only 1/12 scales. Thus, [a/i] is different from the other six pairs in two ways: firstly, there is no immediately apparent phonetic/phonemic reason for its collapse, and secondly a unique contrast,[11] neither language-specific *nor* universal, emerges over its differentiability. Nor does this phenomenon appear to be a product of experimental design.

---

[10] Japanese smallness shares this maximum non-consensus with Spanish passivity, which is the only other scale showing a collapse of [a/i].

[11] This is discounting the [ʌ/i] results (European language speakers 1/12; Asian language speakers 6/12) as skewed by the Japanese cluster, which *was* due to experimental design problems.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

General conclusions on the consensus data. The most immediately striking case is the widespread non-consensus on [ʌ/u]; it is considered likely to indicate an interlinguistic (negative) agreement that there are features in the vowels of these two stimuli causing them to be confused phonosymbolically. I have suggested these are the centrality and acoustic muddiness of [ʌ] and the low frequency concentration of acoustic energy and backness of [u]; intuitions of their similarity may be a genuine language universal.

The above interpretation suggests that the Korean cluster of non-consensus on [ʌ/u], though remarkably large, may not in fact be due to experimentally-caused difficulties with pronouncing an exotic, American English vowel. On the other hand, the Japanese cluster on [ʌ/i] and the Spanish small knot on [ʌ/a] probably *are* due to difficulties in assigning an equivalent first-language vowel to [ʌ]; these findings can therefore be considered as resulting from the experimental design.

Since it is found on two unrelated dimensions, the English knot on [u/a] is most probably a case of language-specific phonosymbolism; however, there is some evidence from the Korean data that the size evaluations for this stimulus pair are shared, indicating a possible universal. The Spanish small knot on [u/i] on the size scales can be traced back to a problem affecting all the Spanish size data, due to phonetic and lexical interference and to the stimulus spellings. Consequently, although language-specific, it is held to be caused by experimental design. On the other hand, the Japanese small knot on [u/i], while most probably also language-specific (confined to one dimension: size, more pronounced on the salient scale, part of a unique configuration on the size rankings) appears not to be a product of experimental design.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Finally, the Korean and Japanese knots on [a/i] can be seen as components of a special case: they are incapable of simple phonetic/phonemic explanation, apparently not produced by experimental design and are neither language-specific nor universal (since unlike [ʌ/u] the nonconensus is not shared by the European language speakers).

To sum up, amongst the twelve unusual instances of non-consensus we have seen instances of universality (e.g., the general collapse of [ʌ/u] across all language groups and dimensions); of unconditioned language-specificity (e.g., the [u/a] collapse on the English energy scales); of experimentally caused language-specificity (e.g. the Japanese non-consenus cluster on [ʌ/i]); and one case of geographically or culturally grouped language-specificity (the Korean-Japanese collapses of [a/i]on brilliance). From table 11.2. we can observe that overall the stimulus pairs are well differentiated. Apart from these twelve unusual instances of non-consensus, there is nothing in the consensus data as organized by stimulus pair which strongly suggests either a language-specific response to the stimuli or a universal one.

11.2.2 Ranking data and universality versus language-specificity

To support the interlinguistic comparison through which issues of the universal versus language-specific bases of phonetic symbolism can be further examined, we require an overview of the ranking data for the four language groups. Since the intent is to allow the emergence of the broad picture, some detail will be omitted. As has been observed in the language-by-language analyses (Chapters 7 through 10), the salient member of each antonymic scale-pair generally yields the same pattern of responses as its non-salient

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

counterpart, but shows that pattern with sharper definition. This observation has its roots in a difference between salient and non-salient scales in the robustness of subjects' judgements for any stimulus pair. For the reader's convenience, Figure 11.1 below brings together the scatter-plots previously presented separately for each language, as a summary means of illustrating the consistency *across languages* of these differences in robustness.

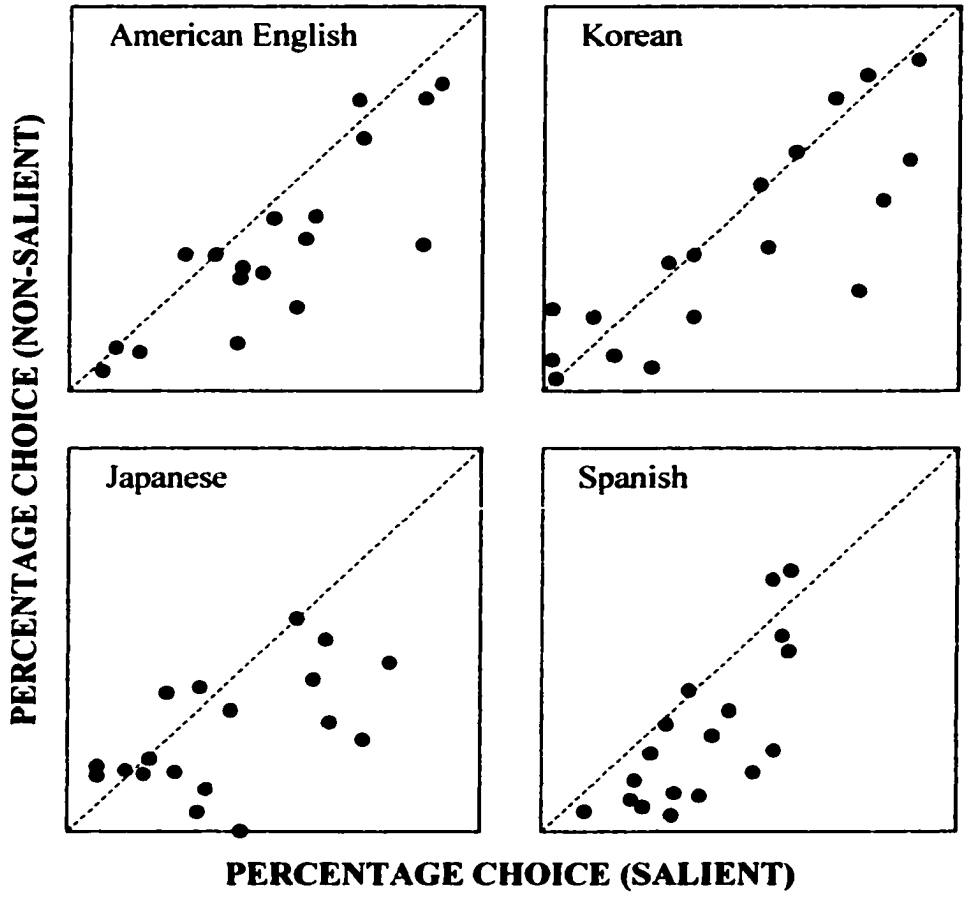
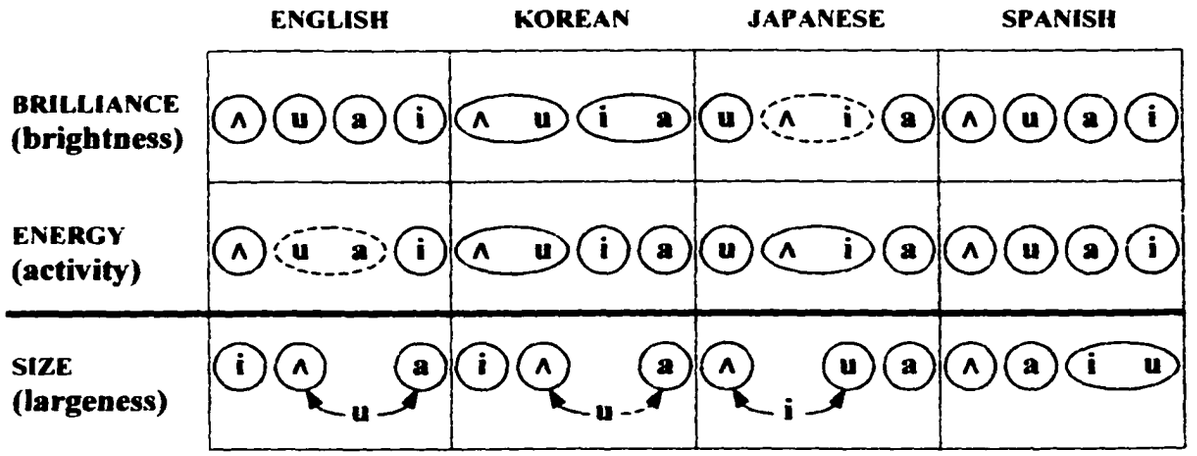


Figure 11.1, Preference strength in antonymic scales across four languages. Percentage of elective choices ('neither' responses omitted) favoring preferred stimuli for each of 6 stimulus pairs evaluated on salient vs. non-salient sub-scales of 3 dimensions. Axis values run 50 to 100%.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The discussion that follows will focus on the more sharply defined data patterns arising with the salient scales. [12] Therefore in Table 11.4 below, summarizing ranking data on the brilliance, energy and size dimensions for the four language groups, only the salient scale of each dimension is represented. Moreover, the table treats the data as ordinal rather than interval, i.e., only ranking sequences and not numerical ranking index values are considered.

Table 11.4: Stimulus rank orders and rank steps, for salient scales



In the table, the non-consensus data discussed earlier have been integrated with the rankings to indicate collapse (non-consensus) and differentiation (consensus) in pairwise

[12] The impression cannot be avoided that for all subjects, whatever their language group, the non-salient terms *dark, passive* and *small* were more difficult than their corresponding salient antonyms for making judgments. It is reasonable to expect that this difference may have been more due to differences in the measures themselves than in the objects measured; i.e., had the stimuli been drawn figures, nonlinguistic noises or flavors, the same contrast in distinctness of attributions might well have been found. It has been observed (Dr. Dianne Bradley, in a personal communication) that the salient term comes without presupposition so that the question *How large were the portions?* does not imply that the portions were large, i.e., is a genuinely neutral enquiry. However, this is not true of the non-salient term , so that the question *How small were the portions?* presupposes that they were small, and the issue is what degree of smallness. Thus any of the 18 questionnaire items using non-salient terms may have been violating presuppositions about one or both stimuli, making those items more confusing for subjects to answer.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

comparisons of stimuli. As mentioned earlier, in interpreting the ranking data it is crucial to know whether rank neighbors are endorsed as independent of each other, i.e., whether by statistical test a stimulus at a given rank is significantly differentiated from its rank neighbors. If it is not so differentiated, then it could be claimed that the ranking is no more than accidental. Collapse with ranking sequences is represented in Table 11.4 in two ways. For straightforward cases, an ellipse surrounding a pair of stimuli indicates that there was no statistical support for their differentiation. For more intricate cases in which a single stimulus enters into two collapses simultaneously—with both its higher- and lower-ranked neighbors, themselves differentiated from each other—arrows link the collapsed stimuli. (On the salient scales, there was no collapse extending over a stimulus triple.) In either case, dashed lines indicate statistical support for differentiation only at the level of trend.

Two terms will be used in discussing the data patterns: *rank order* (or ranking) refers strictly to the ordering of stimuli as dictated by ranking index values; *rank step* takes into account statistically endorsed differentiation (or lack of it) between stimulus pairs. Thus, although the number of stimuli rank-ordered is always four, the number of rank steps may be fewer, since a step may be occupied by two collapsed stimuli; e.g., on the brilliance scale, English and Spanish have four rank steps, while Korean has two and Japanese three. Circles have been added around single stimuli in the table's data display to make the number of rank steps in each sequence clearer. Note that there can be *indeterminacy* in the assignment of stimuli to rank steps, e.g., on the English size scale, [i], [ʌ] and [a] each occcupy a separate rank step, but since [u] is differentiated neither

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from [a] on the highest step nor from [ʌ] on the middle step, its placement is indeterminate.

Given their similar factor-loadings, it was expected that bright/dark might evoke a response pattern similar to that of large/small and different from that of active/passive.[13] The most obvious fact which emerges from Table 11.4 is that, for each of the languages, there is remarkable consistency in the rank orders emerging on the brilliance and energy scales. All four groups agree on the fact of similarity in the brilliance/energy patternings; for any language group, any difference between the scales is restricted to collapse or differentiation with respect to the rank steps. Therefore, as a cover term for both dimensions, the metadimension of *vitality* is proposed, with its constituent scales *vital* (bright, active) and *nonvital* (dark, passive). How language groups construe the set of vowels with respect to vitality may vary, but the evidence for the existence of the metadimension itself is strong. Table 11.4 also indicates an interlinguistic agreement that size is a different dimension from vitality: for each of the languages, there is a remarkable inconsistency between the rankings for vitality and those of size. It is again true that while all four language groups agree that vitality and size do not pattern in the same way, how these patterns differ depends on the language. In what follows, then, how the languages treat the vowel stimuli is reviewed in detail, for the vitality and size dimensions separately.

[13] Yet in this study, large/small proved to be the 'odd one out,' yielding a different response pattern from the other two scales for all four sets of speakers. This is evidence that, at the very least, the factor-loadings of semantic scales are not absolutes but sensitive to whatever is being judged along them.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The vitality dimension. The English and Spanish groups rank the stimuli on the vitality dimension in the order [ʌ]< [u]< [a]< [i], and differ only in that English exhibits some weakness in the middle range (with an [u/a] collapse on activity but not brightness), while Spanish does not. Korean ranks stimuli similarly to English and Spanish, except for reversing rankings in the upper range, so that [a], rather than [i], becomes most-vital. Japanese departs from the other three groups in reversing the rank order of [ʌ] and [u] in the lower range, with [u] uniquely occupying the lowest rank step and [ʌ] and [i] together occupying the middle step. (Note here that [ʌ] and [i] are found on the lowest and highest vitality steps, respectively, for the European language speakers.) In the upper range, the Japanese agree with the Koreans (and disagree with the European language speakers) in ranking [a] above [i]. Arguably, the most important vitality disagreement is between the European and Asian language speakers' rankings of [a] and [i]. Judgments on the most-vital ranking appear to divide the four language groups into two roughly geographico-linguistic (and possibly also cultural) areas. Considering that both Tollman (1887, 1904) and Newman (1933) placed [a] towards the dark/sad/slow and dark ends of their respective vowel scales, the Korean and Japanese results here are surprising.[14]

The size dimension. The pattern of agreement among speaker groups for the size dimension is different from that for vitality; English and Korean speakers agree, while the rankings of Japanese and Spanish speakers differ from these and from each other. It may be relevant that agreement has emerged for languages with large vowel systems (English and Korean, cf. the restricted, five-vowel systems of Japanese and Spanish).

[14] Recall that Tollman's and Newman's investigations involved a much larger range of vowels in English (see Chapter X).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Three rank steps emerge in the English and Korean size judgments: [i]< [ʌ]< [a], with the placement of [u] indeterminate between the middle and uppermost steps. Japanese maintains the English-Korean rank order for [ʌ] and [u] and [a], but places these on separate rank steps; [i], uniquely lowest for English and Korean, is now indeterminate between the lowest and middle steps. Spanish is different again: whereas English, Korean and Japanese place [a] on the uppermost step, Spanish has removed it to the middle range, and whereas English, Korean and Japanese place [i] in the lower range, Spanish promotes it to the uppermost step. Viewing the distribution of collapses and indeterminacies on the size scales, it appears that judgments of non-largeness were easier to make for the English, Korean and Spanish speakers, while judgments of largeness were easier for the Japanese.

Contrasts between vitality and size. Interlinguistic comparison between the scales indicates agreement that vitality evoked better differentiation than size. There are comparatively fewer vitality scale collapses overall, so that rank step indeterminacies are found on the size scales only. This suggests that in general the vitality dimension allowed easier phonosymbolic evaluation of the stimulus vowels than the size dimension.

Differences between vitality and size rankings are most economically captured by observing that for each group, a single stimulus dramatically changes its position in the ranking from one dimension to the other, while there is no change in the relative order of the remaining three stimuli. In the English case, [i] travels from one extreme to the other; ranked as highest for vitality, it is ranked lowest for size. The Koreans also move [i]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

downward from mid-range/ high mid-range for vitality to the lowest step for size, leaving [a] stable on the upper step for both Korean dimensions. Japanese removes [u] upward from the lowest step on the vitality scale to the middle of three steps on the size scale. As with the Korean results, [a] remains stable on the upper step for both dimensions, but in contrast with the Korean scales it is consistently independently ranked. Spanish also moves [u] upwards from a midlow vitality ranking (third of four steps) to a high size ranking, where it joins [i]; the stability and independent ranking of [ʌ] on the lowest step of both dimensions is also a striking feature of the Spanish results.

Triggers of phonosymbolic vitality. For the European language speakers, vitality appears to be determined by three features which in normal situations are interlocked: 1) the location of greatest concentration of acoustic energy of the stimulus vowel; 2) its articulatory correlate of anteriority; and 3) its (articulatory and visual) lip gesture.[15] Here, the three features will be treated as a gestalt, termed the *frequency* gestalt.

The stimulus with the highest vitality ranking is [i] whose vowel has its major concentration of acoustic energy in the high frequencies and the most anterior tongue articulation. It also has lip spreading, i.e., a lip gesture correlating in terms of social interactions with the smile; smiling is associated with brightness metaphorically in both

[15] These cover three perceptual fields: the aural, the proprioceptive/kinetic and the visual. Specially targeted research (such as Eberhardt's (1940), study with deaf subjects) would be necessary to determine the degree to which each of these features may have contributed separately to vitality judgments. Moreover, individual differences in speakers' perceptual focus (more aural, more proprioceptively aware, more visual) may have tended to highlight different vitality features. In terms of further research, this suggests that prior to evaluations, the perceptual focus preference (if any) of each subject should be established.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Spanish and English,[16] and in addition has the acoustic effect of raising the formants (see Tartter, 1980).[17] Stimuli are ranked lower for vitality as their vowels have a lower area of concentration of acoustic energy and a greater degree of backness (in the sequence [i], [a], [u], the tongue position recedes: see Newman 1933). The English [u/a] collapse on the activity scale can be argued to be the result of a *feature equivalence* in which two levels of vitality as judged on the frequency gestalt have been collapsed, i.e., there is an equation between low and middle concentrations of acoustic energy and between back and mid-back tongue positions. Finally, lip gesture on all three point vowel stimuli may have played its part[18] and is certainly expected to have done so for [i] and [u]. Like [i]'s lip spreading, [u]'s rounding has also both a social correlate in the pout, which has an opposite meaning from the smile, and a definite acoustic effect: it lowers the formants (see Borden, Harris and Raphael, 1994).

Lowest-ranked by English and Spanish speakers is the stimulus [ʌ], whose vowel has *centrality features*: a neutral tongue position, a diffusion, rather than concentration of acoustic energy, and no visible lip gesture. Since [ʌ] and [u] occupy separate steps on the European language speakers' vitality scales, it appears for both groups that [ʌ]'s absence of an area of acoustic energy concentration and of definite tongue and lip gestures is experienced as distinctly less-vital than [u]'s low-frequency concentration, backness and lip rounding.

---

[16] Thus, Spanish has, like English, widely-used metaphorical expressions, such as *una sonrisa brilliante*, 'a bright smile,' and so on.

[17] Tartter also notes that lip spreading raises amplitude (cf. the discussion of the intensity gestalt below).

[18] The openness/low jaw position of [a] may have been judged less vital than the lip spreading of [i] but more vital than the lip-rounding gesture for [u].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In contrast, for Korean speakers, [ʌ] and [u] are collapsed on the lowest step of both scales, suggesting that in the lower range there is an equivalence of vitality features operating for the Koreans, i.e., a phonosymbolic equation of [ʌ]'s centrality features with [u]'s (low) frequency-gestalt features. Another contrast between the Koreans and the European language speakers is in the rankings of [a] and [i]. Where English and Spanish speakers rank [i] as most vital, for Korean speakers this rank is taken by [a]. Thus, it would appear that for the Koreans certain features specific to [a] are able to counter—or even dominate—[i]'s high frequency concentration, anteriority and lip spreading. Feature equivalence is suggested by the collapsed rankings of [a] and [i] on the Korean brightness scale; on this group's activity scale, where the rankings of [a] and [i] are independent, domination of the vitality gestalt by one or more new triggering features is indicated.

The most plausible claimants are [a]'s greater perceived loudness (i.e., its lip and jaw gesture), which is linked to its openness. These two features also form a gestalt. Since, as with the frequency gestalt, it is not possible in these data to separate the phonosymbolic effects of openness and loudness, they will therefore be grouped together under the term *intensity* gestalt.[19]

Given that [a] is independently ranked on both Japanese vitality scales, it can be argued from even stronger evidence than in the Korean case that [a]'s openness and loudness are

[19] The effects of the intensity gestalt may have also contributed to the English and Spanish vitality judgments for [a]. However, if intensity were an additional vitality trigger for these groups, it proved to be less potent than the frequency gestalt; this is shown by [a]'s occupying the second (rather than highest) vitality step.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

responsible for its most-vital ranking. Moreover, it can be claimed that contrasting features of [u] relative to the intensity gestalt are responsible for its independent ranking on the lowest steps of both Japanese vitality scales. It is [u]'s proposed marked rounding (i.e., a small-mouth gesture not an intrinsic feature of the Japanese high back vowel [ɯ], but produced especially by these subjects to differentiate their [ʌ] from [u]) and its resulting diminution of acoustic intensity (perceived softness) which oppose it to [a].[20]

As with the ranking of [a] and [u], an explanation involving the intensity gestalt can be found for the collapse of [ʌ] and [i] and their assignment to the middle step for Japanese speakers. In contrast with the stimuli occupying the highest and lowest steps, this pair have in common that they lack a definite lip gesture.[21] It can also be claimed that they are most likely of middle-range acoustic intensity, perceived to be neither as loud as [a] nor as soft as [u].[22] Anteriority (which I have argued that [i] and [ʌ] share, in the Japanese renditions of these stimuli) appears not to have had vitality connotations for this group—or perhaps it had a small, secondary effect, since the (presumably) less-anterior [ʌ] is ranked as less vital than the (presumably) more-anterior [i]. In sum, the Japanese vitality attributions were made on a consistently different basis from those of the European language speakers and can be plausibly explained as triggered by the intensity gestalt.

---

[20] Recall from the discussion of non-consensus data that [u] and [a] are always differentiated by Japanese speakers, which is evidence in favor of a general Japanese [u/a] phonosymbolic opposition.

[21] Japanese does not have lip spreading on [i].

[22] Acoustic intensity has a mid-range value for [i] and [ʌ] for two different reasons: [i] despite its closeness has a high frequency concentration which gives it carrying power (another form of aural saliency) unlike [u]; and [ʌ] though not as open as [a], is not a close vowel and therefore is somewhat loud.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Returning now to the Korean results, we see that this group seems to have participated in two phonosymbolic ordering systems, sharing most of its rankings (and presumeably its criteria for obtaining them) with English and Spanish, but agreeing with Japanese to a large extent on the highest-ranked stimulus, and therefore judging it by a different criterion.[23] A question still to be answered is why Korean speakers distinguished [a] in this way. Since the Korean-Japanese agreement on [a] spans two languages which are not phonetically related, it suggests a response which is not language-specific; however, since it contrasts with the European language speakers' agreement on the most-vital ranking of [i], it cannot be described as universal either.

One possible explanation is that the Asian language speakers' ranking of [a] may be related to a cultural aspect common to Korean and Japanese speech styles, implying different connotations to open-mouthedness from those made by English and Spanish speakers.[24] Another explanation is that the European language speakers interpreted the vitality terms as meaning "excitement-bearing" and connected this with high-frequency noise, whereas the Asian language speakers interpreted the terms as "asertiveness-bearing" and connected this with loudness. In this context, the [i/a] collapse on Korean brightness may be significant; it may indicate an unresolved conflict between judging [i] and [a] according to excitability/high-frequency criteria (perhaps closer to a universal measure of brilliance)[25] and according to assertiveness/loudness (which may be a

---

[23] Korean and Japanese are not in complete agreement on step assignment, since there is the Korean [i/a] collapse on brightness.

[24] Such a hypothesis would require data on the role of mouth-openness/loudness in speech styles in Korean and Japanese in order to support it. Speech style data have been requested but so far are not available.

[25] The hypothesis behind this is that nonlinguistic sounds, such as those of instruments, with concentrations of acoustic energy in the higher or lower frequencies (e.g. the timbre of the harp compared to that of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

criterion culturally shared with Japanese speakers and perhaps other speakers of Asian languages).

Examination of the vitality scales in Table 11.4 indicates that, for all speakers, [ʌ] and [u] are the nonvital stimuli, while [a] and [i] are the vital ones. Although the rank order within either pair will vary as we look from group to group, in no case is either [ʌ] or [u] placed at the vital end of the scales, or [i] or [a] at the nonvital end. Pairing the stimuli in this way will describe responses to both the frequency and intensity gestalten, though rather less effectively for the English energy scale or for either of the Japanese scales.[26] I would suggest that this is due to the aural saliency of both high frequency noise and high intensity noise: judged by either gestalt, [ʌ] and [u] will be non-salient. Grouping by aural saliency would tend very broadly to categorize [i]'s high pitch with [a]'s loudness;[27] whichever of the features then receive the greater weight would depend on the subject group making judgments, and this in turn would decide the ranking of one stimulus above the other.

In further support of this kind of broad categorization, we note that [a] and [i] share nonlow concentration of greatest acoustic energy, i.e., can be seen as roughly located in

---

double bass) will be universally judged with the higher-frequency instruments as having the brighter sounds. This is in the area of sound symbolism, rather than phonetic symbolism.

[26] There is a mid-range collapse of [u/a] on the English energy scale, but the nonvital [u] below vital [a] rank order is maintained, and these stimuli are ranked independently in the same order on the English brilliance scale. On both Japanese vitality scales there is an [ʌ/i] collapse in the mid range (i.e., a nonvital with a vital stimulus), due to language-specific features of the two stimulus vowels' pronunciations, but the rank order of [ʌ] below [i] is maintained.

[27] Association between the two types of aural saliency may be one of the mechanisms behind the operation of feature equivalences.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the same area on a frequency gradient, away from [u] and [ʌ].[28] Moreover, in languages where lip spreading is a feature of [i], the wide mouth opening may have resulted in increased acoustic intensity (see Tartter, 1980), drawing [i] despite its close jaw position nearer to [a] in perceived loudness. At some very general level, then, phonosymbolic judgments produced by the frequency and intensity gestalten agree with each other; this fact can be captured by pairing the stimuli.

Triggers of phonosymbolic size. As observed, the size data are less clearcut than those of vitality; the greater number of collapses and the indeterminacies suggest that potentially competing triggering features may have determined size attributions. Three such features are proposed: *articulatory spaciousness* (anteriority, openness), *acoustic/proprioceptive duration* (perceived length)[29] and *acoustic intensity* (perceived loudness). Spaciousness and duration have a relation to size which is phonosymbolic and iconic (i.e., they are phonetic and gestural), while intensity is related to size sound-symbolically (large objects/creatures tend to make louder noises than do small ones).

Unlike the features making up the two vitality gestalten, the size features can function independently of each other: they can either harmonize, enhancing each other's effects, or conflict, attenuating or even neutralizing each other's effects. In addition, they may give rise to felt equivalences (i.e., a combination of size features of one stimulus may be

[28] Of course, [a], having its greatest frequency concentration in the middle range can also be classified as nonhigh along with [u], which may account for the [u/a] knot of non-consensus in the English data. This does not invalidate the argument associating [a] with [i], but it indicates that [a] may have a double alliegance.

[29] See Sapir (1927, 1929) and Newman (1933) for a discussion of spaciousness and duration triggers.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

equated with a different combination of another stimulus). For all groups except the Spanish speakers, spaciousness appears to be the primary determinant of the rank orders, with less weight given to duration[30] and less again to intensity. The effects of the two acoustic triggers are seen on the distribution of stimuli over the rank steps; what weight (relative to spaciousness) is accorded to one or both will vary depending on the group responding. In the Spanish case, because of subjects' equation of the term "large" with Spanish *largo* 'long,' perceived duration was licensed to become the primary trigger, and spaciousness is claimed to have had little or no influence. Absence of competing triggers is proposed as the reason why there is only one collapse on the Spanish size scale.

Here it is important to remind the reader that not all of the phonetic or phonemic details of these features, particularly those relating to perceived length and perceived loudness, could be supplied for the equivalent vowels in all four languages (see the relevant sections, Chapters 8-10), and indeed reliably large-scale acoustic studies of many aspects of vowels may not yet exist. It is also important to remember that the three size-triggering features are of different orders in relation to the phonologies of the languages. Acoustic intensity is phonetic but does not produce meaningful phonological distinctions, while articulatory spaciousness, though phonetically based, does produce such distinctions (anteriority, height). Finally, perceived vowel duration can be interpreted a) as a purely phonetic feature of the vowel judged, b) as determined by the phonological-phonotactic function of that vowel in the speaker's first language, or c) as the product of the speaker's

[30] The exception is the Japanese ranking of [ʌ], where the weighting of spaciousness and duration appear to have been reversed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

first language phonology interpreting the vowel as it occurs in a second language (in this case, English).

In examining the English-Korean rank order of stimuli, we see the phonosymbolic effects of the (related but independent) vertical and horizontal dimensions of articulatory spaciousness. Some advantage is given to the vertical dimension: i.e., the jaw openness/low tongue position of [a] ranks it as largest, while ranked below it as next-largest is the horizontally spacious [u] with its extended-mouth-chamber lip rounding and back tongue position.[31] This stimulus order agrees with the ranking of [a] above [u] on the large end of size scale given by Newman (1933)—who incidentally ranks [ɔ], which has both vertical *and* horizontal spaciousness, above [u] as the largest vowel.[32] In contrast, [i] is nonspacious both vertically and horizontally, and is ranked by English and Korean speakers as least-large; this vowel is also placed at the small end of Newman's scale. In the middle to lower range, [ʌ] lacks marked spaciousness on either dimension and is ranked as less large than [u] but larger than [i].

Next, we consider the English-Korean agreements on the distribution of the stimuli over the three rank steps of the scale (i.e., the collapses of [ʌ/u] and [u/a], and the independence of [i]). The phonosymbolic effects of the other two proposed size triggers must now be taken into account, namely the extent to which they harmonize or conflict with the effects of spaciousness. We note that [a] has feature harmony, its greater length

[31] In some cases, [ʌ] may also have had vertical spaciousness, depending on how low it was articulated (see Mackay 1987) and [a] horizontal spaciousness, depending on how far back it was articulated.

[32] Tolman (1887, 1904) also ranks [ɔ] at the lowest (largest) end of his scale.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and loudness being phonetic reflexes of its greater openness; it contrasts with [u], which has horizontal spaciousness, but like [i] has a close vowel further muted by rounding. Thus, [u] carries a conflict between its spaciousness (which allies it with the largeness of [a]) and its perceived softness (which allies it with the smallness of [i]): this is responsible for its indeterminacy.

The [u/a] collapse can be explained as due to a felt equivalence between the features of horizontal and vertical spaciousness, that is to say an equivalence between the two subtypes of the dominant size feature. But the conflict hypothesis suggests that [u]'s perceived softness was also exerting a pull in the opposite direction, away from the equivalently spacious but louder [a], and towards [ʌ] and the non-large end of the scale. The [ʌ/u] collapse can be explained therefore as due to an equivalence of the combinations of each vowel's size features. Lacking marked spaciousness on either dimension, [ʌ] is of middling intensity (neither as open and loud as [a], nor as close and soft as the high vowel stimuli); this was felt as equivalent to [u]'s conflicting combination of spaciousness with low acoustic intensity.

Like [a] but placed at the other end of the scale, [i] has feature harmony on two of its size triggers: it lacks both vertical and horizontal spaciousness, and phonetically has the shortest intrinsic duration of the point vowels in AE (see Peterson and Lehiste, 1960).[33] As a close vowel, it can also be claimed as nonloud, though its high frequency concentration may tend to modify the impression of softness by giving it a different type of acoustic saliency. But whereas the feature-harmonious [a] has spaciousness, the most

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

important size trigger, in common with [u], [i] does not share its nonspaciousness with any other stimulus vowel: for this reason, it is independently ranked, in contrast with [a].

Turning to the Spanish data, perceived duration becomes crucial in explaining the rankings; here it must be restated that impressions of vowel duration can be arrived at via more than one feature. Although [a], for reasons of its greater jaw openness, is phonetically the longest of the stimulus vowels, perceived longer duration can also be seen as a phonemically/phonotactically-generated property of the tense, diphthongized open-syllable English vowels; these include [iy] and [uw] *but not [a]*. It has been argued that due to circumstances peculiar to the experiment (English-based experimental design features interacting with native phonosymbolism, cognate confusion and so on),[34] the Spanish group focused on this English-specific feature as the dominant size trigger, and that it affected their perceptions of the stimuli [i] and [u], contrasting them in duration with the (shorter) nonhigh stimuli. Focus on tenseness/diphthongization thus caused the [u/i] collapse and the pair's placement on the highest rank step: the stimulus vowels were judged equally long, therefore indistinguishably large. This further resulted in the paradox that [a] was reclassified as short, ranked immediately above the lowest-positioned [ʌ]; by durational criteria, both stimuli became non-large.[35]

---

[33] Korean vowel duration features are not available.

[34] For full discussion, see Chapter 10, sections 10.2.2 and 10.2.4: mistranslation of "large" as *largo* (long), misreading of the stimulus spellings as indicating two checked and two free syllables, misinterpretation of the AE diphthongized high vowels as being longer.

[35] Recall that [a] also has claims to largeness on grounds of intensity and spaciousness.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Finally, despite agreement between the English, Korean and Japanese speakers in their rankings of [i], [u] and [a], there are still four questions to be answered in accounting for the Japanese data. Why does [a] occupy the highest rank step, independently, in contrast to its collapse with [u] on the highest step of the English and Korean scales? Why is [ʌ]—rather than [i], as on the English and Korean scales—ranked least-large? Why, if [u] and [ʌ] are, as previously claimed, allophonically related for Japanese subjects (and also suffer from a general weakness in differentiability) are they *not* collapsed? And why is [i] indeterminate between the lowest and middle steps?

As in the English and Korean data, the features of horizontal and vertical spaciousness will explain the Japanese rankings of [a] and [u] as largest and second-largest stimuli. But [a] is also the *loudest* stimulus, and in this it contrasts much more noticeably with the rounded [u] for the Japanese speakers than for the other groups (for language-specific reasons; see the discussion above of the Japanese vitality results). It can therefore be claimed that while [u]'s spaciousness ranks it next to [a], its contrasting intensity differentiates it and places it on a lower step.

Contrasting with [a]'s spaciousness, the shared nonspaciousness of [ʌ] and [i] explains their collapsed position at the non-large end of the Japanese scale. However, if articulatory spaciousness were the only size trigger, the rankings should have been reversed, with [i] (as somewhat more anterior than [ʌ]) ranked least-large. Since the two stimuli are not differentiated, one possibility is that their rankings are truly accidental: they are considered of equal height and anteriority, and therefore of equivalent smallness. But there is also a plausible explanation for the ranking of [ʌ] as smallest, based on

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

perceived duration; this is due to Japanese-specific shortness features of the stimulus vowel [ʌ], who associated it with their own particularly collapsible (devoiceable), epenthesizable, shortest vowel [ɯ].[36] Thus, ranked as least-large because it is considered the shortest stimulus, [ʌ]'s lesser duration would partially dominate [i]'s greater anteriority as a size trigger—to the extent of ordering the rankings, but not strongly enough to place the two stimuli on separate steps.

The differentiation, rather than collapse, of [u] and [ʌ] can be explained in two ways. The first supposes that even if their durations were perceived as similar, [u]'s marked feature of horizontal spaciousness for Japanese speakers[37] simply dominated, setting it apart from [ʌ] as larger. But it is also possible that [u]'s marked feature of lip rounding (for this group) may have given it a somewhat longer duration, distinguishing it from the notably short [ʌ].[38] Thus, either [u]'s spaciousness alone was the differentiating size feature, or (if [u] *was* distinguished from [ʌ] by its greater duration), [u]'s features of spaciousness and length enhanced each other.

The indeterminacy of [i] on the Japanese size dimension can be accounted for quite simply. It is due on one hand to [i]'s nonspaciousness, which collapses it with [ʌ], pulling both down towards the non-large end of the scale. On the other hand, [i]'s somewhat

---

[36] For a full account of the Japanese pronunciation of [ʌ] and the shortness of Japanese [ɯ], see Chapter 9, Section 9.1.4.

[37] See again 9.1.4, where it is argued that [u]'s rounding and backing—resulting in oral-chamber lengthening—are claimed as marked features for this group.

[38] The discussion here is necessarily speculative, since no durational values for the Japanese [ɯ]-variant, backed, rounded [u], could be located. But the inherent duration of [uw] is longer than that of [iy] which is longer than that of the lax vowels in English (Peterson and Lehiste, 1960).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

longer intrinsic duration and greater resistance to devoicing[39] than [ʌ] pull it up towards the mid range, collapsing it with [u]. Thus, like the indeterminacy of [u] on the English and Korean size scales, [i]'s indeterminacy can be explained as due to feature conflict: nonspaciousness (connoting smallness) versus nonshortness of duration (connoting mid-range size).

It is less easy to generalize about the size results than those for vitality. Firstly, the Spanish size evaluations of [i] and [a] conflict with those of the three other groups, but even if the Spanish size results are excluded as artefacts of experimental design, what remains is still problematic. One way to characterize the remaining size scales is to say that Japanese opposes the largeness of [a] to the other three stimuli, as having varying degrees of non-largeness, while English and Korean oppose the non-largeness of [i] to the other three stimuli, as having varying degrees of largeness. But then, any generalization over all three scales is lost.

On the other hand, it can be claimed of the English, Korean and Japanese size scales that [i] and [ʌ] tend to be the non-large stimuli (nonspacious and/or shorter in duration), and [u] and [a] the large ones (spacious and/or longer in duration).[40] As in the vitality results, a pairing seems to emerge, but although this fits the ranking facts, it does not account so well for the step distribution of the mid-range stimuli on these three scales, where we find

---

[39] For an account of the greater predisposition of [u] than [i] to devoice in the triggering environment, and how this is probably linked to durational differences, see Chapter 9, Section 9.1.2.

[40] We note that Spanish agrees with the other groups on the non-largeness of [ʌ] and the largeness of [u], although this may be accidental.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

two different collapsed pairs on the middle step of three. In both collapses, there is a non-large stimulus ([ʌ] for the AE group and the Koreans, and [i] for the Japanese) collapsed with the large [u].[41]

Perhaps the best way to generalize about the size scales is to say that once again pairings can be made, but that distinctions between those stimuli falling in the mid range are more blurred than on the vitality scales, due to the effects of the three potentially conflicting features involved in judging this dimension. The extremes of the primary size feature, articulatory spaciousness, are predictable across languages: [a] will be spacious, [i] nonspacious. But with the entry of duration and intensity as potential modifiers of the primary feature, language-specific differences may also enter (namely, the marked shortness of [ʌ] and softness of [u] for the Japanese group). Phonosymbolic size was Sapir's (1927) original focus of enquiry; he thought of it as the most accessible dimension to investigate.[42] But our conclusion would have to be that though size may seem an obvious starting point, the interplay of size features makes it by no means simple to explore. Further enquiry into this dimension cries out for a large sample of stimulus recordings, so that questions about duration and intensity can be anwered objectively rather than speculatively.

---

[41] For the English and Korean speakers, because of the step indeterminacy of [u] there is an implication either that [ʌ] is enlarged (i.e., being pulled upwards towards the large end of the scale) or that [u] is shrunk (pulled downwards). For the Japanese, the step indeterminacy of [i] also implies either that [ʌ] is enlarged or [u], already differentiated from the largest stimulus[a], further reduced in size.

[42] Phonosymbolic size offered "the most likely chance of arriving at relatively tangible results" (Sapir 1929: 226).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Universality versus language-specificity of triggering features. It is proposed that there are two features of the stimulus vowels which trigger attributions of vitality, and three which trigger size attributions:

Vitality

- *Frequency gestalt* (main acoustic energy concentration/anteriority/lip gesture)
- *Intensity gestalt* (acoustic intensity/degree of openness)

Size

- *Spaciousness* (vertical and horizontal)
- *Length* (perceived duration)
- *Loudness* (perceived acoustic intensity)

All of the English and Spanish vitality rankings, and all of the Korean rankings except for that of [a],[43] are attributed to the effects of the frequency gestalt. Since this set of agreements spans two linguistically related and one unrelated language and covers a major part of the data, it seems a clear indication of a universal phonosymbolic response, and therefore of the frequency gestalt as a universal trigger of vitality.[44] The second vitality trigger is the intensity gestalt, which accounts for the Korean and Japanese most-vital ranking of [a],[45] and for all of the remaining Japanese rankings.

---

[43] The Korean [i/a] collapse on brightness suggests the effects of both the frequency *and* intensity gestalten.

[44] The English collapse of [u/a] on activity and the Korean collapse of [ʌ/u] on both vitality scales represent language-specific weakening tendencies which are nonetheless contained inside this general response to the frequency gestalt. Moreover, the orderings of the stimuli in the collapses still agree with the independent rankings on the other three vitality scales.

[45] It may have had a secondary effect in ranking [a] in second place for the European language speakers.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The disagreement between the vitality rankings of Japanese and those of the other three groups (with the exception of the Korean ranking of [a]) can be explained as caused by the Japanese-specific lack of distinguishing lip gestures on the high vowels. This would result in their loci of main concentration of acoustic energy being less polarized than for the other languages with lip spreading and rounding.[46] Thus, due to their greater definition compared to the features of the frequency gestalt, the articulatory openness and acoustic intensity features of the intensity gestalt become the preferred Japanese determinants of vitality. But it is also claimed that the Japanese subjects produced a special case of lip rounding for [u] (to distinguish it from [ʌ]), and that this marked rounding gesture resulted in a polarity with [a] (small mouth versus big mouth, damped intensity versus unobstructed intensity). This suggests that where it can, lip gesture may play an important role in vitality judgments, although it has not been especially emphasized in the literature. It also suggests that in languages where the vowels lack pronounced lip gestures, evaluating them via the intensity gestalt might be the general alternative strategy for arriving at vitality judgments.

On the size dimension, for three out of the four groups tested, the most important determinant in the rankings appears to be articulatory spaciousness, vertical and horizontal. However, two other acoustic triggers, perceived duration and perceived loudness, are hypothesized as interacting with spaciousness to produce feature harmony or feature conflict, duration being the more important of the two. Interactions among the triggers enhance or attenuate phonosymbolic judgments of size, thus affecting step

[46] This remains speculative, as I have not been able to locate any large-scale studies of frequency values for Japanese.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

placements; they are claimed to be responsible for the fact that the findings for size results are more complex and less clear than those for vitality.

It is argued that articulatory spaciousness determines the rank order of [i], [u] and [a] on the English, Korean and Japanese scales, i.e., the major part of the data. Given that these agreements span three unrelated languages, they strongly indicate spaciousness as a universal trigger of size. Agreement on the third-place ranking of [ʌ] in the English and Korean data can also be claimed as due to evaluation on spaciousness criteria, and therefore part of the same universal phonosymbolism. However, the Japanese ranking of [ʌ] as least-large is argued to have been triggered by an unrelated feature, the particularly short perceived duration of the Japanese high unrounded [ɯ] (their [ʌ]-equivalent); this is language-specific.

Articulatory spaciousness is claimed to have played little or no role in triggering the rankings of the Spanish data, with the possible exception of two relations which agree with the other three groups,[47] where it may have had a secondary influence. But perceived duration is hypothesized as the primary influence on all the rankings on the Spanish size scale, triggered by a language-specific construal of aspects of the experimental design by this subject group.

Perceived duration is thus claimed as a trigger for both Japanese and Spanish size evaluations; and since it has an influence in two unrelated language groups, this is

[47] [i] is rank-ordered below [u], and [ʌ] is placed on the rank step below [a].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evidence that it has universal scope. Moreover, long duration as a phonetic correlate of [a]'s open jaw position may also have played some part in the agreement between English, Korean and Japanese speakers' ranking of [a] as the largest stimulus. But since the Spanish and Japanese speakers' duration-related judgments are argued to be the result of language-specific influences, the two instances where these two groups *do* agree with each other (and disagree with English and Korean)[48] should be interpreted as probably accidental, and hence not evidence of universality.

Duration appears to be a universal connoter of size, but whether a vowel is considered as short or long may be based on general phonetic criteria (as in the rankings of [a] by English, Korean and Japanese subjects), or on the phonology/phonotactics of the speaker's first language (as in the Japanese case of [ʌ]), or on the reinterpretation of a second language's phonology by the first (as in the Spanish case of the English high vowel stimuli and the contrast with the nonhigh vowel stimuli). A distinction thus needs to be made between the general tendency to equate an attribute with a segmental feature, and the way that feature is perceived via the sound-structure of a particular language. Though the phonosymbolic trigger may be universal, just which vowel it is applied to may be very much language-specific.

The third size trigger, perceived intensity, has also been claimed to influence judgments. Loudness as another correlate of [a]'s jaw openness, may have contributed to the agreement by English, Korean and Japanese speakers that [a] was the largest stimulus; it is also possible that impressions of softness, due to its closeness, helped in the ranking

[48] Japanese and Spanish agree on the ranking of [ʌ] and the collapse (though not the ranking) of [i/u].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and step placement of [i].[49] But perceived intensity is argued mainly to have affected one stimulus, [u], in conflicting with (and therefore modifying) the effect of its horizontal spaciousness. The perceived softness of [u] has been argued as responsible for its placement on a separate step from [a] on the Japanese scale and for its indeterminacy and collapse with [ʌ] on the second step of the English and Korean scales.

Was there contamination by English? On measures of both vitality and size we see strong ranking agreements between the AE subjects and one other group, and weaker agreements between those two agreeing groups and a third. Thus, there is definite evidence in favor of a universal phonosymbolic response. However, the procedures, questionnaire and target stimulus vowels were all English, and the ESL subjects were almost invariably students of English. Is it possible that the shared knowledge of English exerted a common phonosymbolic influence on all groups? Could some of the cases of interlinguistic agreement be ascribed to an English-based levelling effect on judgments?

On vitality, the agreement is between English and Spanish, and on size between English and Korean. This shift in the ESL group agreeing with AE speakers argues against English as an overall contaminating factor—otherwise, all three ESL groups would be more in agreement with each other and with English in their repsonses. Moreover, there would not be a difference between which group agreed with the English speakers on vitality and which on size, since the degree of English contamination would not be expected to vary with the semantic measure used. Instead, we see that not only are the

---

[49] However, [i]'s high frequency concentration may have produced another form of aural saliency and so modified any effects of low intensity due to its closeness.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

agreeing groups different dependent on the dimension, but the party most diverging from agreement is also different. All of this argues strongly against a general influence of English as a leveller on phonosymbolic responses.

The non-consensus data also suggest language-specific responses, rather than ones dictated by English influence. The most striking example is the English knot of non-consensus on [u/a] on 4/6 scales, covering both vitality and size. This is partially shared by the Koreans, but for size only, and is found in no other group's responses; it therefore appears to be the result of an English-specific tendency which did not affect other subjects. Similarly, the Japanese non-consensus cluster on [ʌ/i] appears quite specific to that language and certainly does not reproduce any tendency in the English data.[50] Moreover, the general picture of non-consensus distribution is different for each language group and does not suggest contamination by English.

It could be claimed, however, that one case of non-consensus, the cross-linguistic weakness in differentiating [ʌ/u] may have been due to English influence. But here, if there was influence, it has been argued as orthographic (the misinterpretability of the spelling *HUH*)[51] and not phonosymbolic. Since the [ʌ/u] collapse is the only case of non-consensus found across all four groups, rather than being a product of leveling, its uniqueness suggests a genuine interlinguistic agreement.

---

[50] See Table 11.2: there is only one other case of non-consensus for this pair (Spanish, smallness).

[51] The use of the letter *U* in the spelling *HUH* may have signaled both [ʌ] and [u], as it can in the spellings *tub* and *tube*, for those readers inexperienced in written English, i.e., the ESL subjects. But the English-speaking subjects, as experienced readers of the language, would have used a whole-word strategy to interpret *HUH* (which is the traditional spelling of the interjection) as [hʌ].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

11.3 Cross-linguistic personalities of the stimulus vowels

Two aspects of a stimulus vowel's personality are important here: its *distinctiveness* (distinguishability from the other vowels) and its *nature*, as defined by the semantic scales. To establish distinguishability, we must look once again at the distribution of nonconsensus and ask how well a vowel tends to be differentiated from any or all of the other three vowels it is paired with in the eighteen questionnaire items in which it occurs: Can it be said cross-linguistically to have a distinct or blurred personality?

In the English and Korean results, [i] is the best differentiated of the vowels and [u] the worst, while [ʌ] and [a] are at about at the same mid-range level of distinctiveness. In contrast, for Japanese, [i] is the worst differentiated, while [a] is the best, closely followed by [u]; [ʌ] alone is in the mid-range. Spanish results suggest that all the vowels have the same degree of distinctness, with [a] and [i] being slightly better differentiated than [u] and [ʌ]. Thus, there is partial cross-linguistic agreement on which vowels have the most distinct personalities, but Japanese and Spanish diverge from it in different ways, Japanese by contradicting the English-Korean order, and Spanish by flattening it out. The honest conclusion must be that four sets of results are not sufficient to answer the question of vowel distinguishability; more language groups should be tested before any claims can be made about this aspect of the vowels' personalities.

To determine the extent of crosslinguistic agreement on the personality types of the vowels, we return to the rankings on the semantic scales, excluding the Spanish size

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evaluations. [52] Taking evaluations vowel by vowel, we see that for English and Spanish speakers, [a] is moderately vital and for Koreans and Japanese strongly vital; English, Korean and Japanese all find it large. English speakers find [i] strongly vital, as do the Spanish speakers, whereas the Asian language speakers find it only moderately vital. The English and Korean groups judge [i] as small; for the Japanese it appears to be of undetermined size, since it is judged as both non-small and non-large. English speakers evaluate [u] as fairly nonvital, as do the Korean and Spanish speakers, while the Japanese find it strongly nonvital. English, Koreans and Japanese find [u] moderately large. English, Korean and Japanese speakers judge [ʌ] as least-vital, while for Japanese speakers it is moderately nonvital. Japanese speakers evaluate [ʌ] as small, English and Koreans speakers as moderately small.

Assuming that the above agreements indicate true universal tendencies in evaluations, we can therefore expect that by a speaker of any language, [a] will tend to be judged as bright, active and large, [i] bright, active and small, [u] dark, passive and large, and [ʌ] dark, passive and small. Once again, we find that the idea of pairing the vowels in terms of their shared attributes and contrasts is useful. In a situation where there is comparison between two vowels considered to share the attribute X, language-specific or possibly cultural factors will then dictate which vowel is judged as the more X of the two. It is in these cases that further research is indicated.

[52] As discussed in Chapter 10 (sections 10.2..2 and 10.2.4), the Spanish size results were made on a different basis from those of the other three groups; they are therefore excluded from that part of the discussion relating to size.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 12

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

12.1 Phonosymbolic triggers, compared with those proposed by Sapir and Newman

The original focus of enquiry of this experiment grew out of the observations of Sapir (1929) who investigated attributions of size, and Newman (1933), attributions of size and brightness, to speech sounds; their findings led them to claim that certain phonetic features acted as attributional triggers. Sapir (1929:235) observes of vowels that the influence(s) triggering size may be "acoustic or kinesthetic or a combination of both." Elaborating on this observation, Newman (1933: 61-62, 68) cites three phonetic factors (anteriority, locus of concentration of greatest acoustic energy, and "increasing size of the oral cavity") as potentially separable; these "work in unison" to rank the front, but not the back, vowels for size, while the first two factors alone are responsible for brightness evaluations. In addition, both Sapir and Newman note apparent anomalies in their size results suggesting that one feature which cannot be linked to the others, perceived duration, also influenced attributions.[1]

The relevant differences between their earlier investigations and this experiment are the inclusions here of a) another semantic measure, the energy dimension, yielding the scales active

[1] See Sapir (1929: 233-5) and Newman (1933: 61-63, 68).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and passive, b) a target central vowel in the stimulus [ʌ],[2] and c) a seventyfive percent majority of ESL- to English-speaking subjects, i.e., three hundred out of a total of four hundred. Both of the earlier investigators included in their studies small numbers of subjects speaking a language unrelated to English, but came to no conclusions about their data.[3]

A number of Sapir's and Newman's findings agreed with individual results of this experiment,[4] and these suggested that in the present investigations of size and brightness the same phonetic influences were responsible for similarities to the earlier subjects' responses. Due to the inclusion of the new dimension of energy, it was discovered that the two pairs of antonymic scales bright/dark and active/passive ranked the stimuli similarly. In many cases, Newman's brightness triggers also 'worked' for energy, suggesting that what was measured in the earlier experiment may not have been so much the semantic dimension of brilliance, but rather a superdimension which could include both energy and brilliance, and which has been termed here *vitality*.

In discussing vitality results, the implications in Sapir's and Newman's remarks that two or more triggering features may exert a combined influence have been taken further: the first two of Newman's factors are considered as inseparable aspects of one vitality trigger, the frequency

---

[2] Sapir and Newman investigated only peripheral vowels.

[3] Sapir included seven Chinese subjects in his study, noting that their responses suggested agreement with his (five hundred) AE subjects, but that many more foreign language participants were needed before coming to any conclusions. Newman included twelve Chinese speakers in his study, but withheld their responses "until further work is done among foreign-speaking individuals" (Newman, 1933: 55).

[4] For example, English and Korean subjects agreed with Sapir's subjects that [i] was the least large/smallest and [a] the largest stimulus. English and Spanish subjects agreed with Newman's subjects that [i] was the brightest/least dark stimulus, and Japanese subjects agreed that [u] was the darkest/least bright.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

gestalt. Lip gesture (especially rounding and spreading, which enhance locus-of-energy-concentration effects) was mentioned as a separate, unreported focus of study by Sapir (see Sapir 1929: 230), but hardly referred to by Newman in his study;[5] however, in this experiment, it is considered as the third member of the frequency gestalt. The concept of the frequency gestalt as a vitality trigger does not contradict any of Sapir's or Newman's conclusions, but it claims a larger semantic scope than that covered by Newman's brightness/darkness triggers. It also asserts that the three phonetic factors creating vitality attributions should not be considered as independently functioning influences, since their effects in isolation from each other have not yet been experimentally analyzed.

On the other hand, a new element is proposed here with the claim that there is a second vitality trigger, the intensity gestalt, which explicitly challenges one of Newman's assertions since he rejects jaw openness as a brightness trigger.[6] Yet Newman's suggested brightness triggers will not account for the finding of this experiment (most clearly seen in the Japanese data) in which the Asian language speaking subjects evaluated [a] as the brightest and most active stimulus; nor will they account for the other Japanese vitality rankings. Here it is proposed that jaw openness, together with its acoustic reflex, greater intensity (which yields perceived loudness), forms an alternative trigger to that of the frequency gestalt. All of the other stimuli on the Japanese vitality scales appear also to be ordered with respect to the intensity gestalt; it is suggested that it is most likely to operate in a language like Japanese with an absence of lip rounding and spreading on its high vowels. It is also suggested that a shared cultural factor may be responsible for the Korean-

[5] See Newman (1933: 62). Lip gesture seems to be included (although not explicitly) in the trigger "size of oral cavity."

[6] See Newman (1933: 68).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Japanese agreement on the vitality of [a], i.e., in the Korean case, another, nonphonetic, trigger is operating, determining application of the intensity gestalt; this type of trigger would be one of a very different nature from any proposed by Sapir or Newman.

Other new phonosymbolic influences were found in investigating the stimulus [ʌ], which was discovered to have a distinct 'personality' despite the apparently neutral nature of its vowel. It tended to be ranked at the low ends of the scales as nonvital, and to a lesser extent also nonlarge;[7] this was taken to indicate that additional vitality and size triggers were formed by its centrality features, which, as normally interlocked, also can be seen to constitute a gestalt. Articulatorily, these are lack of lip gesture and lack of extreme tongue position (a vitality trigger, also interpretable on the size dimension as lack of noticeable articulatory spaciousness). Acoustically, [ʌ]'s centrality means even formant spacing, i.e. acoustic energy diffusion, rather than an area of main concentration (this is a vitality trigger), and since it is neither an open nor a close vowel, middle range acoustic intensity (a size trigger).

In this experiment, articulatory spaciousness was found to be the main size trigger, and this accords with Sapir's (1929: 235) original observations about the largeness of [a] and the smallness of [i]. Moreover, several of the results have been interpreted as giving evidence which supports Sapir's and Newman's conclusions on the influence of perceived duration. The best examples can be found in the strikingly anomalous Spanish size evaluations, where perceived

---

[7] See Table 11.3: English, Korean and Spanish speakers ranked [ʌ] as least-vital. On the size dimension, English and Korean speakers ranked it in the mid to low range while Japanese and Spanish speakers ranked it as least-large. No group ranked it as vital or large.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

duration is argued to be the crucial triggering (and skewing) factor.[8] Where Sapir (1929: 234) observed effects of perceived duration on phonosymbolic size, he noted that "the linguistic (*i.e. phonemic*) factor vs. the merely phonetic one" was in play, thereby alerting the reader to the possibility of language-specific influences on phonosymbolisms. It has been argued that perceived duration may be determined either by language-specific phonological or phonotactic causes, or by universal phonetic causes, and that effects of both can be found in the data of this experiment.

Neither Sapir nor Newman considered the possibility of *feature conflicts or feature equivalences* as influencing evaluations, though such concepts can be seen as coming directly out of their observations on the influence of features on phonosymbolic judgments, especially the conflict between perceived duration and vowel height. In interpreting the data from this experiment, it has been argued that phonosymbolic evaluations of vowels can be based on more complex responses than unified reactions to one or two phonetic features. Where a stimulus is ranked towards the centers of the scales, vowel-internal feature conflict may be the cause, with the relative weights of two conflicting features pulling the stimulus towards both ends of the scales simultaneously.

Thus, in addition to perceived duration, perceived intensity (not mentioned by either Sapir or Newman) has also been proposed as a size trigger, mostly as a means of explaining the ranking and step placement of [u] on the English, Korean and Japanese scales; it is seen as a source of feature conflict, i.e. a brake on largeness evaluations. The ranking of [u] as second-largest (rather

[8] Another instance is the rank ordering of [ʌ] as least large on the Japanese size scales, ascribed as due to its language-specific noticeable shortness of duration for Japanese speakers.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

than largest) stimulus by English, Korean and Japanese speakers and the indeterminacy of [u] on the English and Korean size scales is considered as due to a conflict between its horizontal articulatory spaciousness and its perceived softness (caused by its closeness and small mouth opening).

Feature equivalence is proposed as a felt similarity between the features of two different vowels leading to their nondifferentiation and ranking collapse; it may operate in one of three ways. Firstly, it may equate two related but separate features of two vowels, as in the horizontal and vertical spaciousnesses of [u] and [a], given as the reason for their collapse on the English and Korean size scales. Secondly it may equate ratings on two different gestalten; for example, the collapse of [a] and [i] on the uppermost Korean brightness step is explained as due to an equivalence between [i]'s high score on the frequency gestalt and [a]'s high score on the intensity gestalt. Thirdly, there may be equivalence between the combinations of features of two vowels, as in [u]'s (conflicting) combination of horizontal spaciousness and perceived softness and [ʌ]'s (gestaltish) lack of marked spaciousness combined with middling acoustic intensity. This is argued as responsible for the collapse of [ʌ] with [u] on the English and Korean size scales.

To conclude, findings of the present experiment do not so much suggest conflicts with Sapir's and Newman's list of triggers as they do additions and modifications. Firstly, those triggering features whose effects have not yet been experimentally isolated from each other should be treated as gestalten, and the number of these should be increased: besides the frequency gestalt (whose component elements were mentioned by Sapir and Newman), the intensity and centrality gestalten should be included as vitality triggers.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Secondly, consideration of whatever was prompting the Japanese vitality evaluations and the Korean-Japanese vitality agreement on [a] suggests that there may be cultural predispositions to interpret attributional terms in different ways, leading to a focus on one phonetic trigger rather than another. Since there were so few foreign-language speaking subjects in the earlier experiments, this issue could hardly be explored. However, the checking of all subjects' interpretations of terms (nonphonetic triggers behind the triggers) should be part of any further investigation. Finally, feature conflicts and feature equivalences may play a part in how a vowel is judged; their potential must be taken into account, respectively, to modify an evaluation or to collapse two evaluations together.

12. 2 Summary and suggestions for further research

In the experiment described in the preceding pages, four vowel stimuli were evaluated on three pairs of antonymic semantic scales by means of a questionnaire; the responses of speakers of four languages were gathered and analyzed. The experiment set out to answer three questions. The first was whether a consistent phonosymbolic response could be discovered, as determined by a statistically significant degree of consensus evoked from any or all language groups; it was answered for all groups in the affirmative. The second was to what extent the phonetic symbolism so discovered could be described as universal or language-specific, as determined by agreement on evaluations (rankings on the semantic scales) between the language groups, most especially groups whose languages are not historically related. Results suggested both universal and language-specific tendencies in phonosymbolisms. The third question was about which

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phonetic or phonemic features of the stimulus vowels tested were most likely to have produced rankings and agreements, and a total of five phonosymbolic triggers was proposed. However, in the absence of recordings of the stimuli as pronounced by any of the language groups, no more than hypotheses could be offered in this area.

In the course of the experiment, several design-related problems became apparent. Two major ones were the use of interlanguage vowels by the ESL subjects and the lack of stimulus recordings, while a more minor one was the use of misinterpretable stimulus spellings. Another issue was the very blurred results given by the nonsalient scales. Therefore, one line of further research would be to replicate much of the present study, but with changes to these areas; other useful modifications could include expanding the repertoire of stimuli (i.e. phonemes) tested, and adding to the number of language groups involved.

Firstly, in order to avoid the various types of misreadings considered likely to have occurred in the present study, the visual stimulus signals should be changed in any subsequent investigations. The new signals should be as transparent as possible and involve non-alphabet characters (for example, IPA symbols), screened for any tendency to provoke independent semantic associations in subjects. The need to learn a set of unfamiliar symbols as signals of the stimulus vowels would add time and effort to the experimental task, but the gain in sign-to-stimulus reliability would make the extra outlay worthwhile.

Next, to maximize precision of results, if the same dimensions are used, only the salient

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

antonymic scales should be included, and if more dimensions are added, antonymic terms should be screened for presuppositional asymmetry.[9] Thirdly, it would be helpful to split the original enquiry into a two-level study, with the initial level investigating only first-language vowels, and the later one, second-language vowels; and fourthly, sample recordings of self-produced stimuli should be made at both levels.

A design which separated intralinguistic from interlinguistic evaluations would rule out many of the uncertainties encountered in the present study. The original idea was to obtain uniform pronunciations from the ESL subjects by requiring them to produce American English vowels; in hindsight this was a mistake, since it ran the risk of confusing native and second-language phonosymbolisms. Analogues to the AE point vowels exist in the other three languages tested, but although in any two languages corresponding point vowels may occupy very similar articulatory positions, the vowels will not be identical: they will be differentiated by such language-specific factors as the *base d'articulation*, presence or absence of a tendency to diphthongize, to involve lip gesture, and so on. Thus, asking ESL subjects in this experiment to produce an AE vowel did not guarantee uniform pronunciation by ruling out first-language interference. On the contrary, it ensured that the resulting vowel was likely to be a hybrid and so doubled the number of potential influences, since possible effects both from English *and* the subjects' first language had then to be taken into account.

---

[9] It was observed that the nonsalient scales (darkness, passivity, smallness) yielded considerably more non-consensus and closer ranking index values than the salient ones. This was ascribed as due to the tendency of the nonsalient terms alone to imply presuppositions (for example, *How large?* does not presuppose whether the object to be judged is large or small, whereas *How small?* presupposes some degree of smallness). For a full discussion, see Chapter 11, Section 11.2.2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Another issue involving first- versus second-language vowels was the inclusion of [ʌ], which has no phonemic analogue in Japanese or Spanish, and whose analogue in Korean is a matter of debate. Results from this experiment indicate that central vowels are well worth including in further studies, but, as with the language-specific realizations of [a], [i] and [u], they should be studied initially in subjects in whose first languages they occur. If no analogue exists, for example, to [ʌ] in the subject's first language, the vowel is an exotic, and its evaluation has to be understood as of such, though this does not rule out the possibility that some form of universal phonosymbolism may operate in evaluating it.[10]

In sum, if one wishes to be certain of exploring the phonosymbolism of a language, and not of an English-foreign interlanguage, it is the responses to first-language vowels which must be studied. Thus, in any further investigation, at the first stage, ESL subjects should be asked to produce and evaluate vowels from their own first languages, and given word-lists exemplifying the target vowels and questionnaires in those languages.[11] Naturally, offering the questionnaires in several different languages would raise translation issues, such as the uniformity of reference of the semantic scale terms used across those languages. Like the analogue vowels themselves, maximum similarity of terms, rather than some kind of ideal identity, would have to be pursued, and cultural differences (like the conflicting connotations of passivity for Japanese subjects)[12] taken into account.

---

[10] This indeed is the conclusion about [ʌ] arrived at in the present study (for details, Chapter 11, Section 11.2.1).

[11] Incidentally, this would rule out problems with confusion of cognate words, such as happened in the Spanish case. An intralinguistic study in which Spanish vowels were evaluated with Spanish epithets could answer the question raised by the present experiment: Had there not been the confusion between *largo* and 'large,' might the Spanish and AE size judgments have turned out as close to each other as they did for vitality?

[12] For a discussion of the connotations of 'passivity' in Japanese, see Chapter 9, Section 9.2.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

After an initial, intralinguistic study was made, the basis would be established against which the results of a second, interlinguistic enquiry could be compared; at this stage, subjects would perform the same types of evaluation on second-language vowels.[13] It would now be possible to see how far native phonosymbolisms might affect evaluations of vowels in the second language, and reliable answers on the question of universality could thus be obtained. Let us assume that this second-level study would be, like the present experiment, an enquiry into attributions to American English vowels. What type of stimuli would be used? One possibility is that subjects would listen to and evaluate a set of recorded AE vowel stimuli (culled from a first-level set of tests on AE speakers). This would guarantee stimulus uniformity, but it might produce less interesting results than if the original design were followed, with the different subject groups self-producing stimuli, but in this case also recording them. If, in their pronunciations, the resulting degrees and types of departure from AE analogue vowels could be measured and correlated with phonosemic judgments made, then it would be possible to look at the relation between second-language phonosemic attributions, interference and first-language attributions.

Here it must be repeated that in any further studies of this type, where stimuli are produced by the subjects, it is essential to record the stimulus pronunciations of a reliable sample of each subject group. This would remove from their present realm of speculation observations about which phonetic elements triggered phonosymbolic responses; for example, the extent of the proposed influences of duration and intensity on size judgments could be ascertained by correlating acoustic measurements with judgments evoked. Recording and analyzing the stimuli

[13] For ease of understanding, and to avoid complications due to connotational shifts in terms because of their translation into a second language, questionnaires should remain in the first languages of the subjects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

would add to the time, energy and amount of instrumentation needed in any further study, and some of the original time-economy of having subjects produce their own stimuli would be lost. However, in securing analyzable samples of the objects judged, the gain would be well worth the extra expense.

In interpreting the results of this experiment, two hypotheses were made which could be tested by the inclusion of new language groups in subsequent intralinguistic studies. The first concerns the cause of the Korean-Japanese agreement on [a] as most-vital, rather than [i] as the European language speakers judged it. If, as is hypothesized, some form of areal cultural feature was operating here, then it would be worth focusing on the [a/i] contrast and including for comparison subject groups speaking other Asian languages, and other European-language groups. The second hypothesis concerns the divergence between the Japanese vitality evaluations and those of the other three groups: this was ascribed to the lack of lip gesture in Japanese [i] and [ɯ] compared to those of the high vowels in the other subjects' first languages, and the consequent focus on acoustic intensity/mouth openness (rather than features associated with locus of maximum concentration of acoustic energy) as the trigger of phonosymbolic vitality. To test this, it would be necessary to include other language groups whose high vowels showed a similar lack of lip gesture.

Two further suggestions concern expansion of the range of stimuli. It has been observed that a useful way of describing the results is in terms of pairing the stimuli—but in this experiment there were only four stimuli. Are the pairings simply the product of this aspect of the experimental design? If in a further study, more vowels (for example, the mid peripherals [o] and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

[e], which have analogues in all of the languages tested in this experiment) were added to the stimulus repertoire, then phonosemic pairings or groupings of the vowels could be investigated more fully.

Finally, consonants could also be investigated in terms of their place, manner and position in the syllable. In order to minimize consonantal influence on the vowels tested, the stimuli for this experiment included only the initial, 'colorless' [h]. But once the phonosymbolic identity of a number of vowels had been securely established, then CV and VC stimuli containing those vowels could be evaluated. For example, having evaluated [a], [i] and [u], there is then a basis for testing and comparing [ba], [bi] and [bu], [da], [di] and [du], [ga], [gi] and [gu], and so on.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

VOWEL SOUND QUESTIONNAIRE: PRACTISING THE TARGET SOUNDS

You are asked to pronounce four vowel sounds which you will find in the following four lists of words. Each word will contain an example of the target sound. In some of the word lists, the same target vowel sound will be represented by one or more spellings--don't worry about any spelling differences. Just focus on finding and pronouncing the sound that all the words or phrases in one list will have in common. The target sound is always IN CAPITALS.

Slowly and softly read each list aloud twice, focusing on the sounds represented by CAPITALS.

target sound #1-- HAH

1. POP
2. LOTTery
3. BODy
4. ROCKy
5. SOLid
6. MODelling
7. forGOTTen
8. STOP
9 WATCHing
10. HOLly
11. ROBert
12. PROFit

Have you got the HAH sound now? Then repeat HAH-HAH-HAH-HAH-HAH. If you are not sure, slowly read the list aloud one more time.

target sound #2-- HEE

1. MEETing
2. THREE
3. CLEANer
4. LISa
5. rePEATer
6. EVening
7. PETE
8. TREES
9. aGREEment
10. PEOPle
11. READing
12. poLICE

Have you got the HEE sound now? Then repeat HEE-HEE-HEE-HEE-HEE. If you are not sure, slowly read the list aloud one more time.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

target sound #3-- HOO

1. SHOES
2. LEWis
3. GROOMing
4. unDO
5. GLUE
6. MOVEment
7. SOUP
8. TRUly
9. SUSan
10. ROOTing
11. bamBOO
12 ZOO

Have you got the HOO sound now? Then repeat HOO-HOO-HOO-HOO-HOO. If you are not sure, slowly read the list aloud one more time.

target sound #4-- HUH

(Don't forget: the spellings will be different, but the target sound is always the same. Look for this sound in the UNSTRESSED syllables.)

1. AMuse
2. AFFect
3. ARRest
4. COLLect
5. SUPPly
6. funcTION
7. reacTION
8. SUPPorting
9. soFA
10. cuBA
11. diaNA
12. riTA

Have you got the HUH sound now? Then repeat HUH-HUH-HUH-HUH-HUH. If you are not sure, slowly read the list aloud one more time.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

VOWEL SOUND QUESTIONNAIRE response #

1. In which country or countries did you grow up?

_____

2. If your answer to #1 is not the U.S., how old were you when you came here?_____

3. Sex_____ 4. Age_____

5. First language_____

6. Other languages_____

7. How many years have you been studying English?_____

This questionnaire asks you to make judgements by comparing two vowel sounds with each other. In every question, you are asked to pronounce a pair of vowels and to decide which one seems to you to be the brighter or darker, larger or smaller, more passive or active of the pair. Remember, this is not a test-- there are no wrong answers possible. The important thing is to be honest in your reactions.

Please don't keep checking back to your earlier answers-- you don't have to worry about being consistent in your opinions. Just write down your immediate judgement on the vowel pair in the question that you are answering right now.

Even though you are really trying to make judgements, you may find one or more of the questions especially hard for you to answer. On any question, if you have made a real effort to judge, but neither of the pair of sounds appears to you to be darker or brighter, smaller or larger, more passive or active than the other one, you still have a third option. If you truly can't decide, please write in N for neither.

PLEASE WRITE YOUR ANSWERS IN CAPITAL LETTERS: THEY'RE EASIER TO READ!

1 A1 Make the sounds hah-hoo, hoo-hah, hah-hoo, hoo-hah.
Out of hoo and hah, which one sounds darker?_____

2 B1 Make the sounds huh-hee, hee-huh, huh-hee, hee-huh.
Out of hee and huh, which one sounds smaller?_____

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3 C1 Make the sounds huh-hoo, hoo-huh, huh-hoo, hoo-huh.
Out of hoo and huh, which one sounds more active?_____

4 D1 Make the sounds hah-huh, huh-hah, hah-huh, huh-hah.
Out of huh and hah, which one sounds larger?_____

5 E1 Make the sounds hah-hee, hah-hee, hah-hee, hah-hee.
Out of hah and hee, which one sounds brighter?_____

6 F1 Make the sounds hee-hoo, hoo-hee, hee-hoo, hoo-hee.
Out of hoo and hee, which one sounds more passive?_____

7 A2 Make the sounds hah-hoo, hoo-hah, hah-hoo, hoo-hah.
Out of hah and hoo, which one sounds brighter? _____

8 B2 Make the sounds huh-hee, hee-huh, huh-hee, hee-huh.
Out of hee and huh, which one sounds larger? _____

9 C2 Make the sounds huh-hoo, hoo-huh, huh-hoo, hoo-huh
Out of hoo and huh, which one sounds more passive?_____

10 D2 Make the sounds hah-huh, huh-hah, hah-huh, huh-hah.
Out of huh and hah, which one sounds smaller? _____

11 E2 Make the sounds hah-hee, hee-hah, hah-hee, hee-hah.
Out of hee and hah, which one sounds darker? _____

12 F2 Make the sounds hee-hoo, hoo-hee, hee-hoo, hoo-hee.
Out of hoo and hee, which one sounds more active?_____

13 A3 Make the sounds hah-hoo, hoo-hah, hah-hoo, hoo-hah.
Out of hoo and hah, which one sounds smaller? _____

14 B3 Make the sounds huh-hee, hee-huh, huh-hee, hee-huh.
Out of hee and huh, which one sounds more active? _____

15 C3 Make the sounds huh-hoo, hoo-huh, huh-hoo, hoo-huh.
Out of hoo and huh, which one sounds darker? _____

16 D3 Make the sounds hah-huh, huh-hah, hah-huh, huh-hah.
Out of huh and hah, which one sounds more passive?_____

17 E3 Make the sounds hah-hee, hee-hah, hah-hee, hee-hah.
Out of hee and hah, which one sounds larger?_____

18 F3 Make the sounds hee-hoo, hoo-hee, hee-hoo, hoo-hee.
Out of hoo and hee, which one sounds brighter?_____

19 A4 Make the sounds hah-hoo, hoo-hah, hah-hoo, hoo-hah.
Out of hoo and hah, which one sounds larger?_____

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

20 B4 Make the sounds huh-hee, hee-huh, huh-hee, hee-huh.
Out of hee and huh, which one sounds more passive?_____

21 C4 Make the sounds huh-hoo, hoo-huh, huh-hoo, hoo-huh.
Out of hoo and huh, which one sounds brighter?_____

22 D4 Make the sounds hah-huh, huh-hah, hah-huh, huh-hah.
Out of huh and hah, which one sounds more active?_____

23 E4 Make the sounds hah-hee, hee-hah, hah-hee, hee-hah.
Out of hee and hah, which one sounds smaller?_____

24 F4 Make the sounds hee-hoo, hoo-hee, hee-hoo, hoo-hee.
Out of hoo and hee, which one sounds darker?_____

25 A5 Make the sounds hah-hoo, hoo-hah, hah-hoo, hoo-hah.
Out of hah and hoo, which one sounds more active?_____

26 B5 Make the sounds huh-hee, hee-huh, huh-hee, hee-huh.
Out of hee and huh, which one sounds brighter?_____

27 C5 Make the sounds huh-hoo, hoo-huh, huh-hoo, hoo-huh.
Out of hoo and huh, which one sounds larger?_____

28 D5 Make the sounds hah-huh, huh-hah, hah-huh, huh-hah.
Out of huh and hah, which one sounds darker?_____

29 E5 Make the sounds hah-hee, hee-hah, hah-hee, hee-hah.
Out of hee and hah, which one sounds more passive?_____

30 F5 Make the sounds hee-hoo, hoo-hee, hee-hoo, hoo-hee.
Out of hee and hoo, which one sounds smaller?_____

31 A6 Make the sounds hah-hoo, hoo-hah, hah-hoo, hoo-hah.
Out of hah and hoo, which one sounds more passive?_____

32 B6 Make the sounds huh-hee, hee-huh, huh-hee, hee-huh.
Out of hee and huh, which one sounds darker?_____

33 C6 Make the sounds huh-hoo, hoo-huh, huh-hoo, hoo-huh.
Out of hoo and huh, which one sounds smaller?_____

34 D6 Make the sounds hah-huh, huh-hah, hah-huh, huh-hah.
Out of huh and hah, which one sounds brighter?_____

35 E6 Make the sounds hah-hee, hee-hah, hah-hee, hee-hah.
Out of hee and hah, which one sounds more active?_____

36 F6 Make the sounds hee-hoo, hoo-hee, hee-hoo, hoo-hee.
Out of hoo and hee, which one sounds larger?_____

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

response #

How would you translate the following words into your first language? (You can use more than one word in your translation.)

S-37 SMALL_____

L-38 LARGE_____

A-39 ACTIVE_____

P-40 PASSIVE_____

B-41 BRIGHT_____

D-42 DARK_____

You are asked to write down three things which you normally think of as having the qualities of the words you have just translated.

Ex: What are three things that you normally think of as SOFT?

A FLOWER PETAL, A BABY'S CHEEK, A CLOUD

D-43 1. What are three things that you normally think of as DARK?

_____

B-44 2. What are three things that you normally think of as BRIGHT?

_____

P-45 3. What are three things that you normally think of as PASSIVE?

_____

A-46 4. What are three things that you normally think of as ACTIVE?

_____

L-47 5. What are three things that you normally think of as LARGE?

_____

S-48 6. What are three things that you normally think of as SMALL?

_____

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

response #

How would you translate the following words into your first language? (You can use more than one word in your translation.)

D-37 DARK_____

B-38 BRIGHT_____

P-39 PASSIVE_____

A-40 ACTIVE_____

L-41 LARGE_____

S-42 SMALL_____

You are asked to write down three things which you normally think of as having the qualities of the words you have just translated.

Ex: What are three things that you normally think of as SOFT?

A CLOUD, A BABY'S CHEEK, A FLOWER PETAL

S-43 1. What are three things that you normally think of as SMALL?

_____

L-44 2. What are three things that you normally think of as LARGE?

_____

A-45 3. What are three things that you normally think of as ACTIVE?

_____

P-46 4. What are three things that you normally think of as PASSIVE?

_____

B-47 5. What are three things that you normally think of as BRIGHT?

_____

D-48 6. What are three things that you normally think of as DARK?

_____

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 2: Self-reported expererience in learning/speaking English (ESL subjects)

# of years	Korean	Japanese	Spanish
unspecified	2	4	8
0—11 months	6	3	38
1—2+	3	3	13
3—4+	5	1	6
5—9+	47	51	20
10+	37	38	15

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX 3

Summary of elective choice frequencies and consensus data

Two tables (differently organized) present elective choice frequencies (*'neither'* responses omitted) for paired stimuli judged on antonymic sub-scales of three semantic dimensions. Data order within cells follows the order of the stimulus-pair listing; bolding in cells shows non-significant frequency contrasts. p>.05; dashed cell outlines show trends, .10>p>.05.

A. Presentation by Language Group

	BRILLIANCE		ENERGY		SIZE	
	Brightness	Darkness	Activity	Passivity	Largeness	Smallness
American English speakers						
<i/a>	74 25	27 71	71 29	35 65	36 64	67 32
<i/u>	85 15	12 86	77 21	29 67	26 72	64 34
<i/ʌ>	94 5	10 89	83 14	17 82	32 67	67 32
<a/u>	79 20	27 72	**58 41**	**55 45**	**54 43**	**44 55**
<a/ʌ>	92 7	12 87	90 7	30 67	76 22	38 59
<u/ʌ>	69 29	**43 55**	70 29	35 64	**54 46**	**47 52**
Korean speakers						
<i/a>	**44 56**	**59 40**	23 77	68 31	19 78	81 19
<i/u>	85 15	12 88	63 37	**47 53**	24 76	76 23
<i/ʌ>	88 11	9 91	65 35	33 66	32 68	67 32
<a/u>	93 6	20 80	88 12	37 63	**58 41**	**54 45**
<a/ʌ>	95 5	7 93	89 9	25 74	68 32	**40 59**
<u/ʌ>	**48 51**	**48 51**	**50 48**	**44 54**	**50 48**	39 60
Japanese speakers						
<i/a>	40 60	**59 40**	33 66	**55 44**	37 63	**57 42**
<i/u>	78 18	25 75	69 28	34 65	**46 53**	**41 58**
<i/ʌ>	**59 40**	**42 57**	**57 43**	**42 58**	**53 46**	**41 56**
<a/u>	89 11	28 72	85 14	38 62	66 34	31 68
<a/ʌ>	77 22	22 77	81 18	35 63	79 20	30 69
<u/ʌ>	38 62	68 32	29 71	**50 50**	65 34	**47 52**
Spanish speakers						
<i/a>	79 20	26 72	66 34	**45 55**	78 22	39 60
<i/u>	77 22	17 82	71 27	33 63	**43 53**	**50 45**
<i/ʌ>	80 20	16 84	65 35	36 64	65 34	**51 47**
<a/u>	60 35	39 59	69 31	**53 44**	32 67	67 31
<a/ʌ>	69 29	37 62	74 24	**41 56**	61 37	**44 50**
<u/ʌ>	59 38	**44 52**	60 38	**42 55**	79 21	24 74

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 3, continued

B. Presentation by Stimulus Pair

	BRILLIANCE		ENERGY		SIZE	
	Brightness	Darkness	Activity	Passivity	Largeness	Smallness
Stimulus Pair <i/a>						
English	74 25	27 71	71 29	35 65	36 64	67 32
Korean	**44 56**	**59 40**	23 77	68 31	19 78	81 19
Japanese	40 60	**59 40**	33 66	**55 44**	37 63	**57 42**
Spanish	79 20	26 72	66 34	**45 55**	78 22	39 60
Stimulus Pair <i/u>						
English	85 15	12 86	77 21	29 67	26 72	64 34
Korean	85 15	12 88	63 37	**47 53**	24 76	76 23
Japanese	78 18	25 75	69 28	34 65	**46 53**	**41 58**
Spanish	77 22	17 82	71 27	33 63	**43 53**	**50 45**
Stimulus Pair <i/ʌ>						
English	94 5	10 89	83 14	17 82	32 67	67 32
Korean	88 11	9 91	65 35	33 66	32 68	67 32
Japanese	**59 40**	**42 57**	**57 43**	**42 58**	**53 46**	**41 56**
Spanish	80 20	16 84	65 35	36 64	65 34	**51 47**
Stimulus Pair <a/u>						
English	79 20	27 72	**58 41**	**55 45**	**54 43**	**44 55**
Korean	93 6	20 80	88 12	37 63	**58 41**	**54 45**
Japanese	89 11	28 72	85 14	38 62	66 34	31 68
Spanish	60 35	39 59	69 31	**53 44**	32 67	67 31
Stimulus Pair <a/ʌ>						
English	92 7	12 87	90 7	30 67	76 22	38 59
Korean	95 5	7 93	89 9	25 74	68 32	**40 59**
Japanese	77 22	22 77	81 18	35 63	79 20	30 69
Spanish	69 29	37 62	74 24	**41 56**	61 37	**44 50**
Stimulus Pair <u/ʌ>						
English	69 29	**43 55**	70 29	35 64	**54 46**	**47 52**
Korean	**48 51**	**48 51**	**50 48**	**44 54**	**50 48**	39 60
Japanese	38 62	68 32	29 71	**50 50**	65 34	**47 52**
Spanish	59 38	**44 52**	60 38	**42 55**	79 21	24 74

Refer to Chapter 6 (Section 6.3, *Data Types and Treatment*) for details about the way that the frequency data shown here relate to subjects' judgment behavior, and for details about the analysis underlyling the statistical claims indicated here with distinctive outlines and bolding.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX 3

Summary of ranking index values

The table presents ranking index values calculated (within language groups) from the frequency data provided by subjects' responses to stimuli presented as pairs. Values are calculated for each stimulus, separately, by pooling over sets of inter-related judgments made on any sub-scale of a semantic dimension; when considered relationally, these provide a succinct characterization of behavior with respect to that scale. For discussion of the way in which the ranking index metric is formulated to serve (in conjunction with pairwise statistical comparisons) as a device for distilling data patterns from raw elective choices frequencies, refer to Chapter 6 (Section 6.3, *Data Types and Treatment*).

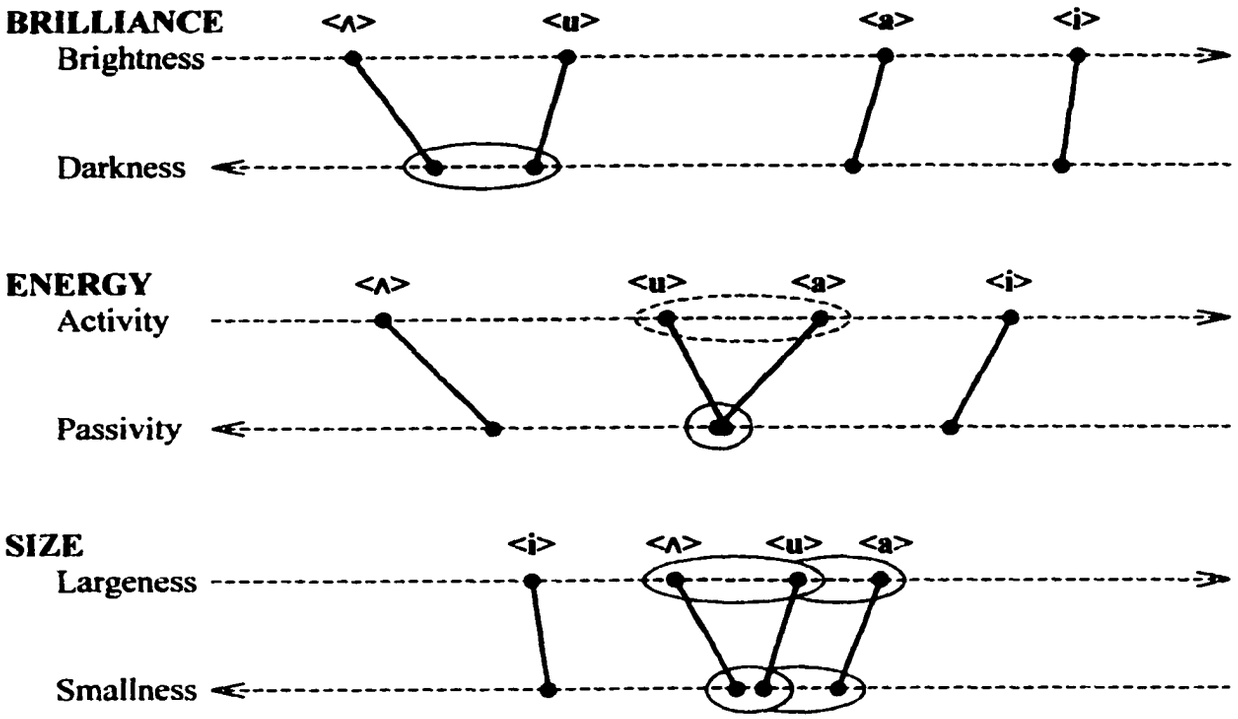
	BRILLIANCE		**ENERGY**		**SIZE**	
	Brightness	**Darkness**	**Activity**	**Passivity**	**Largeness**	**Smallness**
American English speakers						
<i>	.849	.166	.783	.275	.316	.669
<a>	.660	.372	.598	.505	.658	.386
<u>	.350	.681	.446	.498	.573	.459
<ʌ>	.139	.780	.171	.722	.455	.485
Korean speakers						
<i>	.726	.268	.503	.497	.253	.752
<a>	.816	.224	.852	.312	.689	.379
<u>	.232	.722	.332	.537	.562	.360
<ʌ>	.225	.786	.311	.655	.497	.508
Japanese speakers						
<i>	.600	.423	.537	.440	.456	.471
<a>	.758	.303	.781	.394	.696	.347
<u>	.226	.717	.240	.592	.510	.582
<ʌ>	.416	.557	.441	.574	.337	.600
Spanish speakers						
<i>	.792	.199	.678	.385	.631	.479
<a>	.510	.502	.594	.507	.387	.588
<u>	.399	.631	.399	.514	.675	.344
<ʌ>	.295	.671	.328	.595	.310	.590

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 4

Data patterns for four language groups

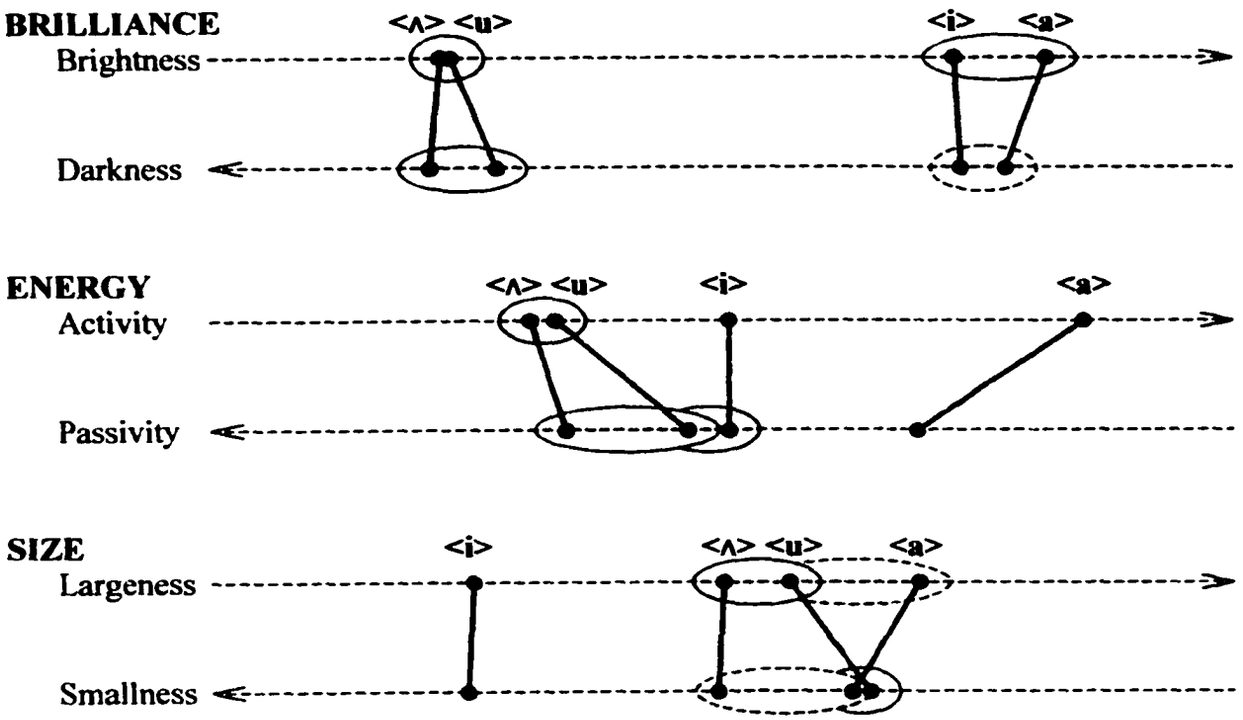
AMERICAN ENGLISH SPEAKERS: Ranking index values and consensus data for salient and non-salient sub-scales of three semantic dimensions. Ellipses surround stimulus pairs not differing by statistical test, $p>.05$; dashed outlines indicate instances of trend, $.10>p>.05$.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 4, continued

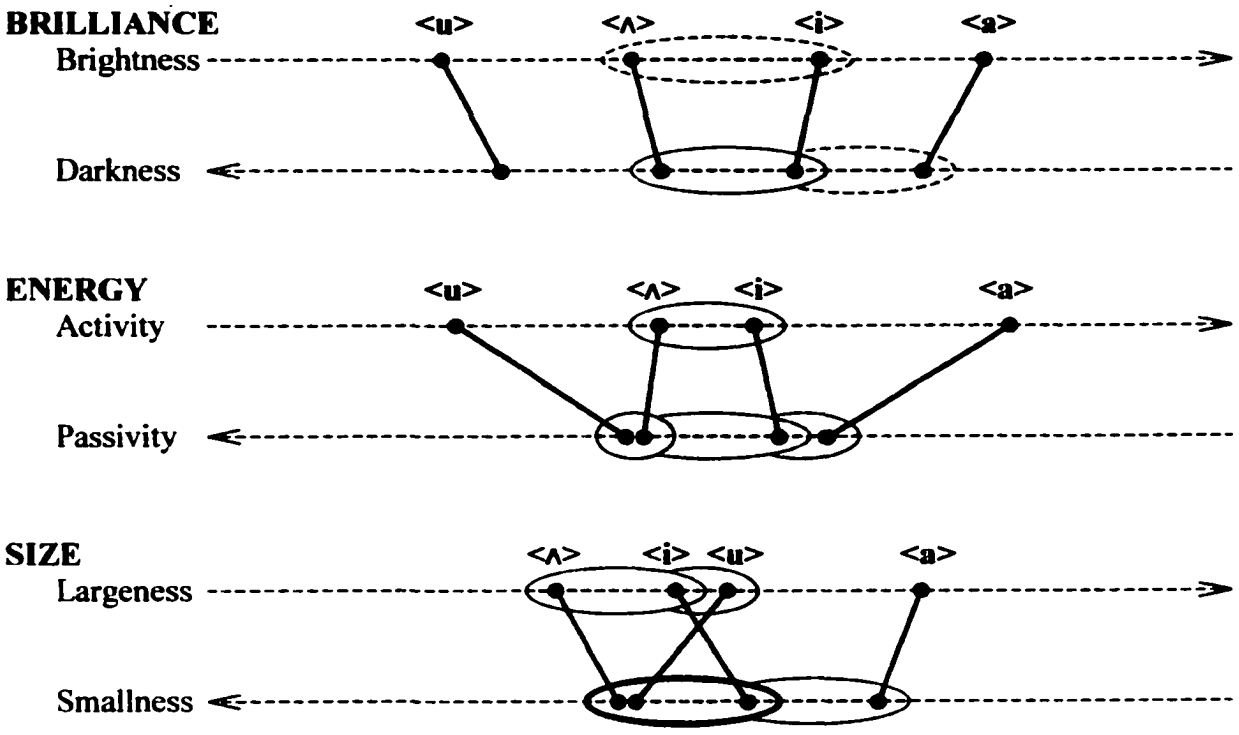
KOREAN SPEAKERS: Ranking index values and consensus data for salient and non-salient sub-scales of three semantic dimensions. Ellipses surround stimulus pairs not differing by statistical test, $p>.05$; dashed outlines indicate instances of trend, $.10>p>.05$.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 4, continued

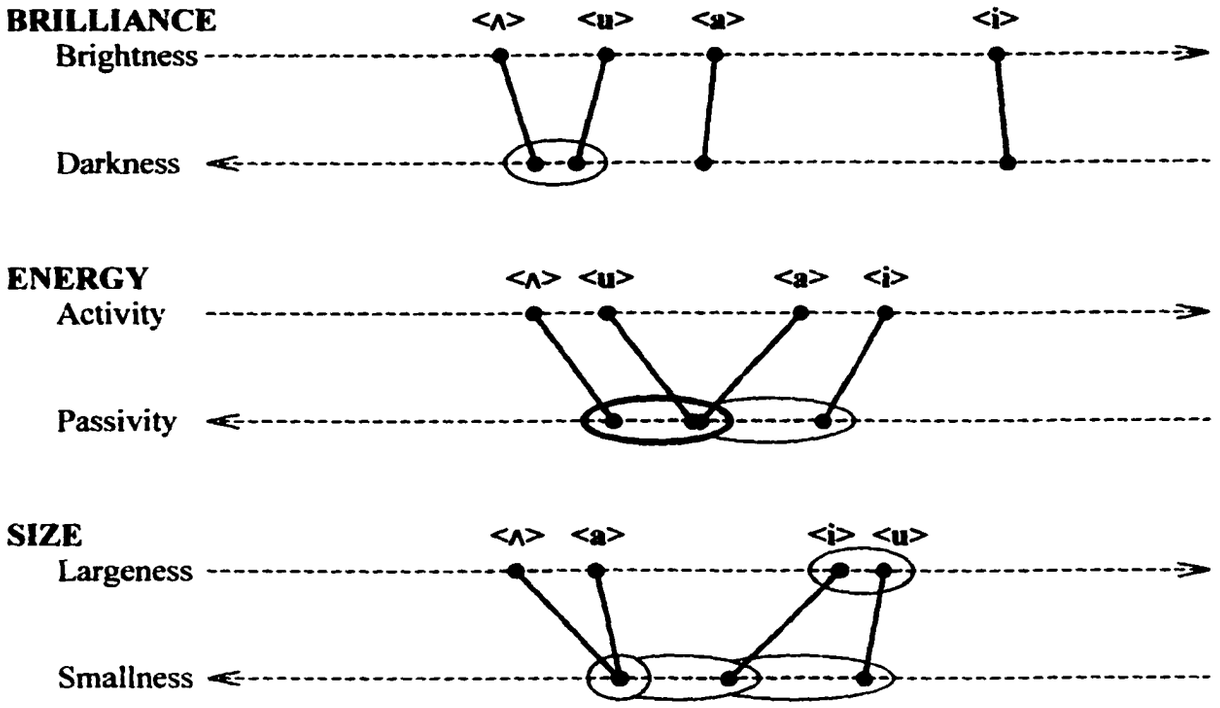
JAPANESE SPEAKERS: Ranking index values and consensus data for salient and non-salient sub-scales of three semantic dimensions. Ellipses surround stimulus pairs not differing by statistical test, $p>.05$; dashed outlines indicate instances of trend, $.10>p>.05$.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 4, continued

SPANISH SPEAKERS: Ranking index values and consensus data for salient and non-salient sub-scales of three semantic dimensions. Ellipses surround stimulus pairs not differing by statistical test, p>.05; bolding indicates a special case with no significant contrasts among a set of *three* stimuli.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 5

Elective choice frequency matrices for four language groups

AMERICAN ENGLISH SPEAKERS: Elective choice frequencies for paired stimuli on salient and non-salient sub-scales of three semantic dimensions ('neither' responses omitted). Shading picks out cells with non-significant contrasts, p>.05; dashed lines indicate instances of trend, .10>p>.05.

BRILLIANCE

Brightness

	<u>	<a>	<i>
<ʌ>	69 / 29	92 / 7	94 / 5
<u>		79 / 20	85 / 15
<a>			74 / 25

Darkness

	<u>	<a>	<i>
<ʌ>	43 / 55	12 / 87	10 / 89
<u>		27 / 72	12 / 86
<a>			27 / 71

ENERGY

Activity

	<u>	<a>	<i>
<ʌ>	70 / 29	90 / 7	83 / 14
<u>		58 / 41	77 / 21
<a>			71 / 29

Passivity

	<u>	<a>	<i>
<ʌ>	35 / 64	30 / 67	17 / 82
<u>		55 / 45	29 / 67
<a>			35 / 65

SIZE

Largeness

	<u>	<a>	<i>
<ʌ>	54 / 46	76 / 22	32 / 67
<u>		54 / 43	26 / 72
<a>			36 / 64

Smallness

	<u>	<a>	<i>
<ʌ>	47 / 52	38 / 59	67 / 32
<u>		44 / 55	64 / 34
<a>			67 / 32

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 5, continued

KOREAN SPEAKERS: Elective choice frequencies for paired stimuli on salient and non-salient subscales of three semantic dimensions ('neither' responses omitted). Shading picks out cells with non-significant contrasts, $p>.05$; dashed lines indicate instances of trend, $.10>p>.05$.

BRILLIANCE

Brightness

	<u>	<a>	<i>
<ʌ>	48 / 51	95 / 5	88 / 11
<u>		93 / 6	85 / 15
<a>			44 / 56

Darkness

	<u>	<a>	<i>
<ʌ>	48 / 51	7 / 93	9 / 91
<u>		20 / 80	12 / 88
<a>			59 / 40

ENERGY

Activity

	<u>	<a>	<i>
<ʌ>	50 / 48	89 / 9	65 / 35
<u>		88 / 12	63 / 37
<a>			23 / 77

Passivity

	<u>	<a>	<i>
<ʌ>	44 / 54	25 / 74	33 / 66
<u>		37 / 63	47 / 53
<a>			68 / 31

SIZE

Largeness

	<u>	<a>	<i>
<ʌ>	50 / 48	68 / 32	32 / 68
<u>		58 / 41	24 / 76
<a>			19 / 78

Smallness

	<u>	<a>	<i>
<ʌ>	39 / 60	40 / 59	67 / 32
<u>		54 / 45	76 / 23
<a>			81 / 19

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 5, continued

JAPANESE SPEAKERS: Elective choice frequencies for paired stimuli on salient and non-salient subscales of three semantic dimensions ('neither' responses omitted). Shading picks out cells with non-significant contrasts, $p>.05$; dashed lines indicate instances of trend, $.10>p>.05$.

BRILLIANCE

Brightness

	<u>	<a>	<i>
<ʌ>	38 / 62	77 / 22	59 / 40
<u>		89 / 11	78 / 18
<a>			40 / 60

Darkness

	<u>	<a>	<i>
<ʌ>	68 / 32	22 / 77	42 / 57
<u>		28 / 72	25 / 75
<a>			59 / 40

ENERGY

Activity

	<u>	<a>	<i>
<ʌ>	29 / 71	81 / 18	57 / 43
<u>		85 / 14	69 / 28
<a>			33 / 66

Passivity

	<u>	<a>	<i>
<ʌ>	50 / 50	35 / 63	42 / 58
<u>		38 / 62	65 / 34
<a>			55 / 44

SIZE

Largeness

	<u>	<a>	<i>
<ʌ>	65 / 34	79 / 20	53 / 46
<u>		66 / 34	46 / 53
<a>			37 / 63

Smallness

	<u>	<a>	<i>
<ʌ>	47 / 52	30 / 69	41 / 56
<u>		31 / 68	41 / 58
<a>			57 / 42

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 5, continued

SPANISH SPEAKERS: Elective choice frequencies for paired stimuli on salient and non-salient subscales of three semantic dimensions ('neither' responses omitted). Shading picks out cells with non-significant contrasts, p>.05. (The data show no instances of trend, .10>p>.05.)

BRILLIANCE

Brightness

	<u>	<a>	<i>
<ʌ>	59 / 38	69 / 29	80 / 20
<u>		60 / 35	77 / 22
<a>			79 / 20

Darkness

	<u>	<a>	<i>
<ʌ>	44 / 52	37 / 62	16 / 84
<u>		39 / 59	17 / 82
<a>			26 / 72

ENERGY

Activity

	<u>	<a>	<i>
<ʌ>	60 / 38	74 / 24	65 / 35
<u>		69 / 31	71 / 27
<a>			66 / 34

Passivity

	<u>	<a>	<i>
<ʌ>	42 / 55	41 / 56	36 / 64
<u>		53 / 44	33 / 63
<a>			45 / 55

SIZE

Largeness

	<u>	<a>	<i>
<ʌ>	79 / 21	61 / 37	65 / 34
<u>		32 / 67	43 / 53
<a>			78 / 22

Smallness

	<u>	<a>	<i>
<ʌ>	24 / 74	44 / 50	51 / 47
<u>		67 / 31	50 / 45
<a>			39 / 60

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix 6

Fónagy's thinking on phonetic symbolism

Ivan Fónagy published two important books on phonetic symbolism—in his term "phonetic metaphor"—from the perspective of one who was both a Darwinian[1] and a Freudian. The earlier, and for the purposes of this dissertation, more central, book, *La métaphore en phonétique* (1979), surveyed much theoretical (phonetic, psychological, stylistic, semiotic) and experimental writing on the subject; it also attempted to answer several crucial questions arising from the surveys.

The first was why grammarians appear to have always *needed* the descriptive resource of phonetic metaphor. Fónagy presents material showing that from the earliest times (i.e., from descriptions of Sanskrit onwards) and in describing a wide number of languages both Indo- and non-Indo-European (such as Arabic and Chinese), grammarians have found it necessary to resort to metaphoric characterizations of speech sounds in order to supplement their more technical descriptions of pronunciations. A second question, growing out of this, was how the phonosymbolic attributions of both the premodern grammarians and humans in general (especially children) demonstrate that they "know" what they apparently do not consciously know. In explaining this paradox, Fónagy ends his book with a discussion of the nature of metaphor itself, and of the dynamic psychological structures that give rise to it.

---

[1] Darwin's importance is more emphasized in the later book, which treats emotional expressivity, but *The expression of the emotions in man and the animals* is cited here too. The centrality of expressive gesture to phonetic metaphor -- of the principles of serviceable associated habits and of antithesis--is never in doubt.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Fónagy's initial chapter surveys the use of phonetic metaphor in early to nineteenth-century grammars and traces the cross-linguistic prevalence of certain symbolic descriptions, such as the thinness of /i/, the comparative fatness of the back vowels and the association of labialization with femininity and anteriority. He notes that vowel height also corresponds to antonymic symbolism, but that such symbolisms vary more from language to language (Fónagy 1979: 4-11). In his later book, *La vive voix* (1983) Fónagy again treats the metaphorical use both of the back-front and the high-low dimensions of vowels. He notes that there is a historical tendency in many languages to move vowels forward; the fronting of vowels is generally associated with good taste and backing with vulgarity. In a test using 20 children and 20 adults, he found that /u/ "smelt bad," while /i/ "had a nice smell" (Fónagy 1983: 84). Vowels necessitating "a vertical movement, notably a large degree of opening, can equally trigger an unfavorable judgement" (Fónagy 1983: 84) because of associations between the mouth and the vagina.[2]

Observing that his own two children (aged 4 and 7) found it easy to make symbolic dstinctions between two speech sounds,[3] he then reports on a series of experiments using young children of pre-reading age (3-5), high schoolers and adults. In his later book, these experiments were again discussed. All three groups agreed with each other on their evaluations, but the adults and young children showed higher consensus in their responses. Fónagy attributed the lower consensus of the high schoolers to the initial impact of orthography—together with normal adolescent

---

[2] Fonagy's basic framework of discussion is strongly Freudian; speech sounds, and later, voice qualities are categorized as anal, urethral or genital; for example apical, trilled /r/ (presumeably a Hungarian phoneme) is described as phallic, while /u/, on account of its retraction and closeness to the pharynx has a "sadico-anal" quality (Fonagy 1983: 104), hence its identity as ugly, bad-smelling and so on.

[3] The capacity of children to respond phonosymbolically (because they are less separated from their physical perceptions of the sounds they make and hear) is one of the underlying themes of the book.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

contrariness. This study used pairs of individual speech sounds as stimuli, the task being to judge which member was the larger, darker, more bitter and so forth. His initial subjects were Hungarian; later he tested adult speakers of English and French, noting that for the subjects, the sounds evaluated "hardly seemed to change character in passing the frontiers" (Fónagy 1979: 54).[4]

From the point of view of vowel symbolism, most striking amongst his results are a number of tables, in which contrasts between /i/ and /u/ are clearly demarcated, and in the expected way, /i/ being smaller, thinner, more agile, less dark and so on. As a follower of Darwin, Fónagy tends towards universalism; he refers comparatively little to language-specific influences on phonetic metaphor, although a thread of non-universal observations can be traced.[5] Yet here he takes care to note that while /u/ was considered more bitter by French and Hungarian subjects, English speakers considered /i/ more bitter, due he feels, to the actual stimulus-word "bitter."[6]

When commenting on certain gaps of agreement in judgements, he remarks that there is a part of "purely subjective" association in phonetic symbolism which "comes from an individual, isolated experience... On the other hand, the diversity of colors, images and ideas evoked can be explained by the *inherent diversity* in such and such a sound which links different articulatory,

---

[4] This of course raises the question of how far the stimulus sounds can be said to be the same when heard/spoken by different language speakers.

[5] For example, he notes that Arab grammarians have certain vowel symbolisms which are the opposite of almost every other language mentioned. Fonagy relates this to the extreme backness of the back consonants of Arabic: a piece of evidence that individual phonologies of languages also have their phonosymbolic part to play.

[6] Intuitively, I agree with his English subjects; there may indeed be language-specific reasons for the difference, but the vowel in "bitter" is actually /ɪ/ not /i/, which however *is* found, confusingly enough, in "sweet."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

acoustic and functional[7] phenomena" (Fónagy 1979: 49, emphasis added). This point is made again in the later book: "Each sound of the language (which one perceives for functional reasons as a simple, homogeneous function) is in reality a bundle of physiological and acoustic features. Consequently, it can lend itself to the representation of different pulsional whims following one or other of its distinctive features" (Fónagy 1983: 103-104). These observations weaken his universalist position but agree with Werner and Kaplan's concept of the "plusrisignificance of vehicular material" (Werner and Kaplan 1963: 218-219).

As befits someone who dedicated *La métaphore en phonétique* to Jakobson, Fónagy includes chapters on phonosymbolic studies in both linguistics and poetry. There is an extensive survey of phonetic symbolism experimentation to date;[8] he comes to the conclusion that the minimal pair technique, using in his term "logatomes" (artificial words), is the most successful at tapping into phonosymbolic intuitions: "Only tests done with logatomes allow one to directly infer the preverbal onomatopoetic (phonosymbolic) competence of subjects" (Fónagy 1979: 53).

In contrast, he finds the word-matching technique dubious, since it "depends very clearly on the presence of onomatopoetic tendencies in today's languages... Positive or negative results obtained by this method underline the presence or absence (of these tendencies)... It is risky to draw conclusions about the symbolic value of sounds starting out with words in a foreign language *without taking into account the phonetic structure of those words*" (Fónagy 1979: 53, emphasis added). Moreover, such studies "have the inconvenience of not separating the

[7] That is, language-specific, *phonological* function.

[8] He includes reports of work by Hungarian, Italian and French researchers not mentioned anywhere else that I have

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

onomatopoetic sensitivity of the subject from objective onomatopoetic tendencies in the lexicon" (Fónagy 1979: 54). Finally, asking subjects to create their own phonosymbolic words, in Fónagy's opinion, "unites the inconvenience of both others (i.e., minimal pair and word-matching methods) without offering any advantage" (Fónagy 1979: 54). Subjects are working with idiosyncratic, non-social language, and there is no way of controlling the influence exerted by the first language and other languages spoken.

He cites certain very specific vowel findings which are worth noting. Ertel (1969) used the semantic differential with German subjects and found that potency scales gave higher scores for /a/ and /o/ (as contrasted with /i/, /e/ and /ey/). Activity scales reversed the scores, with /i/ more exciting than /a/: "Labialization diminishes excitation; in other words, F_2 covaries with degree of excitation and F_1 with the potency values.... Short vowels are more exciting than longer ones" (Fónagy 1979: 55). Fischer-Jørgensen (Perceptual Dimensions, 1967) found a positive relation between F_2 frequency and thin/fat, bright/dark and hard/soft and F_1 frequency and small/big, narrow/wide, taut/relaxed and shut/open (Fónagy 1979: 64).

Following his chapter on phonetic symbolism studies in poetry, Fónagy lists a wide range of traditional associations of attributes with speech sounds, dividing them into those which are perception-related and those judgement-related. A summary of the vowel symbolisms is very much as one would expect, but perhaps worth including.[9] An interesting but speculative section

---

found, and a long list of Soviet experiments (in Russian).

[9] Under the general heading of PERCEPTION comes touch-- *hard sounds* = closed vowels, such as /i/ and /y/ ; sight-- *bright sounds* = front vowels (versus back vowels); touch and sight-- *small sounds* = closed vowels such as /i/ and /y/ (versus open vowels, such as /a/) and front vowels (versus back ones); *thin sounds* = closed/front vowels (versus

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

follows on the possible bases of these associations,[10] and a discussion of a second series of experiments again using paired individual speech sounds in which stimuli and questions were the same as before, but the subjects were 20 deaf children and 20 blind children. These experiments were further discussed in *La vive voix* (1983).

Fónagy compared results with those of the previous series in order to see how far acoustic and visual cues were responsible for triggering symbolic associations. Generally, the deaf children agreed with the hearing, but there was less agreement on such attributions as smallness, brightness, thinness and gaiety.[11] Thus, deaf and hearing subjects found that /i/ was smaller and brighter than /u/, but the deaf children showed less consensus[12] about the smallness, fatness or agility of /i/ demonstrating that "the acoustic aspect enters into play" (Fónagy 1983: 70). Moreover, for the deaf children, /i/ was stronger than /u/: an opposite result from that of the hearing and blind subjects.[13] He suggests that this may have been triggered by "a greater sensitivity to muscular and tactile sensations" rather than to sounds (Fónagy 1979: 115).

---

open/back ones); *light sounds* = front vowels (versus back ones); *angular sounds* = nonlabialized front vowels (versus labialized front or back ones); *hollow sounds* = velars, such as /u/ and /o/ versus front vowels such as /i/. Under the heading of JUDGEMENT comes dynamic: *agile sounds* = closed front vowels (versus back vowels); *strong sounds* = back vowels; esthetic sensation: *pleasing sounds* = front vowels; attitude: *joyous sounds* = front vowels; sex: *male sounds* = back vowels.

[10] One detail is especially worth noting, since it implies that the functional (i.e. phonological, linguistic) aspect of a speech sound may also determine its symbolic characteristics. Fonagy remarks that French schwa (*e muet*) connotes femininity because of its phonological instability.

[11] However, /i/ was perceived by the deaf to be stronger and less sharp than it was by the hearing, which Fonagy ascribes to cuing off articulatory (lingual) tension rather than high frequencies by the deaf group.

[12] However, 20 is a small sample.

[13] I suspect that English speakers would agree with Fonagy's deaf subjects, not with the hearing. In other words, since the participants in these comparisons between deaf and hearing evaluations were, presumeably, Hungarian

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the later book, speculating on what gives rise to judgements, Fónagy remarks: "There is probably a relation between the closed nature of /i/—the very reduced distance between tongue and palate--and the idea of 'smallness' or 'thinness'" (Fónagy 1983: 72), but since the hearing subjects showed more agreement in their evaluations, he concludes that the high frequencies of /i/ also play a part, noting that there is a relation between /i/'s high frequencies and the voices of small animals, women and children in contrast to the low frequencies of /u/ and associations with vocalizing larger animals and large adult humans (presumeably males).[14] However, he also notes (weakening his universalist position but agreeing with Werner and Kaplan's concept of "plurisignificance of vehicular material") that the salience of different features may be responsible for the different strength-evaluations of /i/ and /u/. Thus "acoustic association" influences hearing[15] subjects to find /u/ stronger, while "lingual tension in the articulation of /i/" makes deaf children perceive /i/ as the stronger vowel (Fónagy 1983: 73).

The blind children in general had the clearest responses, perhaps due to especial auditory sensitiveness; this even included good response to the brightness dimension. Fónagy concluded that although his findings were not definitive, proprioceptive input appeared to be roughly 8 times as important as auditory input, which itself was approximately 4 times as important as visual input. These proportions were calculated by comparing the results of the (pooled) hearing subjects of the previous experiment with those of the blind and the deaf; if both handicapped

---

(their native language was not mentioned) there may also be some language-specific factor in play.

[14] For a discussion of this aspect of phonetic symbolism and its extension into the animal kingdom, see essays by Morton and Ohala in *Sound Symbolism* (1994).

[15] Hungarian, but perhaps not English; see above, fn 6.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

groups showed similar response patterns to the non-handicapped, then Fónagy reasoned that the symbolic triggering must have come from non-visual, non-auditory stimuli.[16]

The final chapter of *La métaphore en phonétique* tackles the central question of how phonetic metaphor can be generated by preconscious or unconscious perceptions. What is it that young children—and premodern grammarians—"know " about articulatory movements and positions that allows them to spontaneously understand phonetic symbolisms like *wet*, *bright* and *hard*? Fónagy observes that to perceive a phonetic segment, especially one whose identity is relative to a string (for example, a long vowel) requires a series of enormously complex cognitive operations; these still cannot be duplicated by machine. The same speakers who can get intonation absolutely correct in performance cannot say if their voice is going up or down: "They get hard of hearing as soon as they try to describe their performance, to bring it up to the surface, crossing over the threshold of consciousness... Only metaphors allow the filtering—all the while deforming them—of certain pieces of information" (Fónagy 1979: 164).

He then proposes two possible models of metaphor. The first uses a "very rough representation" (Fónagy 1979: 165) of three phonetic stimuli as examples, /i/, /o/ and /m/. They are contrasted in terms of their most salient articulatory characteristics: high front tongue position for /i/, lip rounding for /o/ and lip closure for /m/. Fónagy's first model is of a railway or tram system, with phonetic departure points in the unconscious zone and tracks leading first through the unconscious, then the preconscious and finally out to destination points the conscious zone. Paths

[16] He quotes here Jespersen and Paget (1930) in support of the view that articulation has an especially important role (i.e. the dominant one) in the formation of phonetic metaphor and in the expressive functioning of sounds.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to the conscious are potentially direct and straight, but in metaphor they become deviated; single or double bends represent repressions in one or both non-conscious zones.

In Fónagy's words, "These (phonetic) terms directly correspond to features that can be represented. One could then join stimuli and reactions (the perceived features) by a straight line. These same stimuli trigger in other cases verbal reactions which are not in direct relation to the phenomena they designate: 'thin sound,' 'round sound.' In the case of metaphoric ideation, at a certain point, one leaves the straight line. If one considers the straight lines as memory-traces... one could say that in the case of these metaphoric denominations, excitation has been transmitted from one memory-trace to another, and that this shift has taken place in the preconscious for /i/, described as thin, and for /o/ as round; and in the unconscious for /m/, perceived as sugary, honey-sweet" (Fónagy 1979: 165).

It seems that where there is a choice between preconscious metaphoric ideation or conceptual analysis, the former is preferred, even if both are equally available. In the case of his seven-year-old daughter, "at this age she was still perfectly capable of noticing the position of her phonatory organs—an operation which appears to be easier for children than adults, including phoneticians—even so, she took the longer but less onerous route in characterizing sounds by metaphor" (Fónagy 1979: 167). Fónagy emphasizes the "painful nature of abstract ideation and the psychic tension it generates... In the case of semantic evolution, one starts by renouncing the act in order to reduce it to its gestural reproduction. The gestural activity is then reduced to the movement of phonatory organs. Articulatory and sound (acoustic) expression, the signifier which is initially joined to the object by resemblance, detach themselves totally from the object as soon

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

as they become its arbitrary sign. This representation becomes in its turn less and less sensual, more and more 'abstract,' thus 'withdrawn, detached (from its object).' The path of abstract ideation is signposted with abdications, with the renunciation of direct gratification that the object produces" (Fónagy 1979: 167). Thus, he concludes that preconscious ideation is experienced as less costly than conscious mental activity.

Fónagy's second model of phonetic metaphor is more complex, allowing for different, even conflicting, symbolisms. His diagram show a stimulus sound (in his example, /i/: this includes the articulatory-acoustic experience associated with it) as the circular center of two larger, concentric circles. In his own words, "A pebble thrown in a lake and creating circular waves propagated in the water, apparently indefinitely, offers a better model of ideation... which could highlight the diffuse, expansive quality of the metaphoric process. The little central circle represents the stimulus that elicits a number, indefinite in association, of preconscious and unconscious ideas" (Fónagy 1979: 169).

In Fónagy's diagram, a number of rays shoot out from the center, each one representing the way a different acoustic, articulatory (or, presumeably, functional) feature of the sound activates different attributions, lends it different qulities. The same feature may give rise to a number of attributions, some of which are more directly related to the sound, less extended, than others. Thus where each ray bisects the inner concentric circle, an attribute such as *sharp*, *hard*, *small* or *thin* is marked. Where the same ray bisects the outer circle, another attribute, related to the first but possibly more specialized, more applicable, even, to animate referents,[17] is marked. The outer

[17] I am indebted to Dr. John Pittman for this observation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

circle attributes carry some inherent contradictions with each other; for example, *thin* on the inner circle is followed by *agile* on the outer, *little* on the inner circle, by *weak* on the outer. *Weak* and *agile* are not mutually exclusive qualities, but neither are they a good match, since *agile* connotes energy.

Marked on Fónagy's diagram at the attribution-points on the rays are also little dotted sections of other circles: "To preconscious associations are added unconscious images (fantasmes)... the number of preconscious and unconscious associations that are triggered is limitless. The totality of these associations that one cannot and does not even try to disentangle is reflected in our consciousness as a vague, global impression... *If one wants to allow for the complexity of the stimulus, such as a vowel or a consonant, one must admit several circles of irradiation*" (Fónagy 1979: 169—170, emphasis added). These eccentric, dotted circles may represent several different phonosymbolic (metaphoric) activities.

One, using Fónagy's own description, is that the dotted circles may represent other perceptions of the same sound (by the same speaker-hearer at another moment, or by another speaker-hearer) in which *agile* is a primary attribute and *thin* a secondary one. Another is that an attribute such as *agile*, which is secondary, two steps away from the center of /i/ as it were, may be primary for a different sound-stimulus, for example, /ɪ/ while *thin* then becomes an outer-circle, secondary quality. Thus, following this second model of metaphor, the grammarian describes /i/ as bright due to the retention of a particular association evoked by the vowel (and probably also in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

opposition to the contrary impressions of /u/).[18] But when he contrasts /i/ with /ae/, he must retrieve another association, such as sharpness or smallness.

Conceptual thought does the opposite of metaphor: it returns to the *features* of the sound and looks at what is accidental, what is individual, what is conditioned phonetically and what by the performance of a specific utterance. Perhaps, Fónagy speculates, the preliminary phase of conceptual thought is free, spontaneous, non-localized association, perception and understanding being at this stage very diffuse. This kind of mental activity obeys the pleasure-principle, by reducing psychic tension through agreeable solutions; it offers both intellectual economy (since there is less effort expended on repression) and also catharsis (since tensions are released by freely allowing the self to fantasize). Conceptual thought, on the other hand, obeys the reality-principle; it is characterized by selective concentration, using a series of filters that allow the passage of certain facts so as to subject them to a fine-grained analysis. Metaphor occupies thus a mid-point between free, preconscious fantasy and directed, controlled thought: it is *controlled fantasy.*

Yet if the subject in a minimal pair phonetic symbolism test is in effect choosing from a number of preconscious associations, since these are not intrinsically embarrassing or upsetting, why do they not jump spontaneously into her consciousness? The heart of the problem is, in Fónagy's words, "Why is it so difficult or impossible to be aware of what one is doing with one's tongue when pronouncing /i/ or /u/?" (Fónagy 1979: 174). In fact, there is more than one reason for the

---

[18] We see here Darwin's first and second principles working in tandem.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

repression of those immediate phonic sensations, and even of many of the associations that they give rise to. Firstly, since "the frontier between the preconscious and the unconscious is permeable" (Fónagy 1979: 174), free associations are often pulled backwards by the unconscious. Therefore, not every association turns out to be so free of embarrassment; Fónagy's illustration is that of the phallic imagery brought by up by trilled /r/ (which in Hungarian phonology he interprets as the erectile segment par excellence); another is the association of /m/ with sweetness via the unconscious connection with nursing. Better for one's mental stability to repress such imagery.

Moreover, if there is one barrier to be crossed between unconscious and preconscious, there is another between preconscious and conscious: experimentally speaking, this is where a knowledge of phonetics could make a difference in subject responses, because phonetic training is designed to bring articulatory movements back across the threshold of consciousness (in this, it resembles certain forms of body work, Zen training and so on). Fónagy observes that automization lowers the amount of conscious operations we have to perform; this is particularly needed in language, where we have to do an enormous amount very quickly. We tend to have strong resistance to consciously re-atomizing an activity, linguistic or otherwise, which has already been assembled and automated for most of our lives: we are questioning some of the bases of our existence, and for most of us this is not a comfortable activity.

In addition, there is something about the nature of language itself which works against our consciously noticing the sounds of which it is made: "the arbitrary character of verbal signs allows us to deny and ignore the substance of speech sounds. The formulation of de Saussure is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

very characteristic in this regard: 'Moreover it is impossible that a sound, a material element, itself belongs to language' (Cours 1972: 64)" (Fónagy 1979: 174-175). This point is made again in *La vive voix* (1983): our capacity to use phonic material in a symbolizing, metaphorical way depends on the special status of phonemes in the structure of language. Fónagy points out, again quoting de Saussure, that even though it is impossible that sound, the material element, itself belong to language (rather than speech), "this does not... prevent phonic material from existing; on the contrary, its lack of linguistic existence assures the (articulatory) gesture an anonymity which allows the speaking subject to profit fully and enjoy the performance of oral gesture with impunity. The metaphors of phonetic terminology are emanations of this performance, partly preconscious, partly unconscious" (Fónagy 1983: 148).

Finally, according to Fónagy, much of phonic sensation is obscure quite simply because it is happening inside the mouth, and this kind of internal body experience is difficult to monitor. His explanation, very much based on Freudian theory, is that each perception has first to cross the unconscious; this explains both why there are distortions of what is perceived and why perecptions are censored as soon as they pass the threshold of consciousness. Although exteroceptions (such as the visual and tactile experience of a ball held between the fingers) seem to be less distorted, or the distortions engendered can be more easily corrected, interoceptions are not so easily brought into awareness or fixed. To sum up, "multiple diverse causes can unite to make one ignorant of what one could know" (Fónagy 1979: 175).

At this point, we are faced with the other side of the paradox. If such ignorance is the rule, how then are the grammarian, the child, the subject in a phonetic symbolism experiment still able via

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

metaphor to know and name what they are ignorant of? Fónagy suggests that certain other mental phenomena can give us a clue. First he cites the struggle to retrieve a forgotten name. Rather in the manner of the second model of metaphor, the inaccessible (according to Freud, repressed) name acts like a pebble dropped in the water, radiating associations out from the center; as we struggle to remember, other names come up which are either semantically or phonically related.

If one focuses on the signifier, the phonetic correspondances dominate, the inaccessible name being analyzed and discomposed into partial anagrams. The searched-for composing phonemes may undergo changes or distortions: these could be symbolic, for example, replacing one back vowel by another perceived as darker, because of one's (repressed) feeling about the referent. If one searches using lexical stress or number of syllables, one is likely to succeed; if one searches using segments, one is more likely to fail because of the chance of phonic distortions.

However, the forgotten word still continues to act; it attempts to rise to consciousness while suffering all kinds of distortions. The same fate is reserved for repressed childhood memories, which according to Freud can be represented by screen-memories. "Substitution is made on the basis of metaphoric and metonymic analogies (resemblance and contiguity) or oppositional relations (negative screen-memories)...[19] Everything happens as if the eclipsed object plunged into the unconscious could act on our consciousness by irradiation, and as if the perception of these emanations had as its condition the object's eclipse, in the same way that the solar eclipse allows the sun's halo to be observed" (Fónagy 1979: 177).

[19] Once again, we see the working of Darwin's (1955) first and second principles.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Although the central consciousness does not let through a perception of backness or tenseness in a vowel, the original repressed sensations are nevertheless activated and in turn activate metaphoric associations. Here Fónagy emphasizes the importance of the semantic differential as a technique for tapping into hidden phonic sensations. He remarks that it makes systematic use of "semantic irradiation... without having fully become conscious[20] of the psychological basis of the procedure. In other words, it involves making experimental poetry by obliging subjects to participate in a game of forced metaphors. And all subjects, without exception,[21] find not the slightest difficulty in it" (Fónagy 1979: 178).

Another mental phenomenon that may illuminate the "paradox of practical, indirect knowledge of things" is the subliminal perception of visual and auditory stimuli. These too, though not consciously registered, can generate all kinds of associations. Fónagy's example is of a dyslexic subject who could not read the word "butter" but who nevertheless was able to associate this apparently semantically uncomprehended word with others which were meaning-related, such as "milk" and "cow."

Moreover, given the unconsciousness of phonic activity and perception, Fónagy is able to explain the sensorially diffused, *synesthetic* character of phonetic metaphor. Because of the subconscious or preconscious nature of acoustic and articulatory sensations, these sensations do not receive investment of energy from the preconscious: this means that they do not become verbalized.

---

[20] It is not very clear from his wording whether he means the consciousness of those who designed the test or that of the subject; he seems to imply that of Osgood et al.

[21] From my own observation, there was considerable variation in how easy subjects found semantic differential tasks.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Instead, they are elaborated following the laws of primary processes; since there is no intellectual control, representations which are discrete in reality become jumbled together, identified through subjective criteria, impressions, perceptions. Thus stimuli which arrive simultaneously via different sense organs are "treated as equivalent and interchangeable" (Fónagy 1979: 180).

According to Fónagy, phonetic metaphor reflects a projective tendency, since interior sensations are externalized: "The most arbitrary metaphors recapitulate a long chain of associations triggered by a pharyngo-buccal gesture" (Fónagy 1979: 181). For this reason, experiemtnally it is not so efficient to use Jacobson-Halle terms (such as compact/diffuse) to release phonosymbolic intuitions. These terms will not resonate so well with subjects as will the more traditional attributes "because they do not represent typical unconscious universal fantasies" (Fónagy 1979: 181). Being based on spectral analysis, they do not correspond to preconscious or unconscious perceptions.

Metaphorical expression means unpredictability. Fónagy remarks that in metaphor there is "no stable, controllable relation between the signified and signifier" (Fónagy 1979: 181). It operates like Freud's description of a witticism, with an instantaneous jump through a long chain of associations, or in Freud's (quoted) words, "A preconscious thought is released for an instant to unconscious elaboration and the result is immediately seized by the conscious attention" (Fónagy 1979: 182). And for the grammarian (or the conscious self of a mature language speaker) the expression is also arbitrary. The only rule is that the metaphor must be *improper*, that is, unexpected--it must give the metaphorist the freedom to play. An *adequate* term cannot be a

---

It seems most likely that repressions will vary individually.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

metaphor because it will not trigger the crucial associations. In sum, the metaphoric term allows unconscious material to surface for several reasons. Firstly, it is extremely rapid. Secondly, it works discreetly and indirectly, through allusions. Thirdly, it is (apparently) operating under the aegis of linguistic arbitrariness. And finally, it has the jester's license to play.

The last question to be answered by Fónagy is why grammarians and linguists in general, (not to speak of other scientists and philosophers) need to use metaphor to communicate thought. He starts by observing that as soon as he is no longer simply reporting on the metaphors of others but trying to describe the thing itself, he too depends upon metaphoric expression to convey his ideas. Although (in contrast to rational, left-hemispheric thought) right-hemisphere brain activity can be linked with metaphoric expression in its "preferance for image, its global, concrete, sensual, direct nature, its swiftness and ease" (Fónagy 1979: 192), Fónagy asserts that metaphoric activity cannot be so crudely localized. It makes use of both sides of the brain.

As Suzanne Langer had pointed out over thirty years earlier, language in general is marked by metaphor,[22] even where it has apparently outgrown it. "The signified changes without keeping the memory of its former shape; the signifier is gifted with memory. Each sign, verbal, lexical or grammatical, registers, in the form of a new sense, the trace of a recurrent metaphoric use, without necessarily giving up its original meaning. The semantic structure of the sign keeps the traces of semantic excursions, in the form of superimposed meanings that one could compare to

---

[22] For a clear and very convincing description of the whole phenomenon, see Langer (1942: 139-142); her book is mentioned in both Fonagy's bibliographies. Fonagy's thinking here evidently grows out of Langer's general idea that we are always obliged to resort to metaphor when expressing what is new.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the rings on a tree which mark the annual revolutions" (Fónagy 1979: 194-195). Fónagy illustrates this with a description of the surface variety and unlikeness of the "so-called possessive constructions" (in French) that "must be interpreted as diversity at the level of conscious analysis and identity at the level of unconscious 'paleological' interpretation" (Fónagy 1979: 195).

This accumulation or layering of meanings must not be seen as having a diachronic structure—no 'befores,' no 'afters'—but in terms of synchronic polysemy. In artificial languages this is avoided, but polysemy is a normal part of ordinary language use; Fónagy points out that formerly monosemous, precise scientific terms get co-opted into normal language for polysemous use. It seems that we as humans need this in order to communicate ideas. All signs recapitulate past thoughts, and we need this kind of recapitulation: "Psychic evolution has the singular particularity of preserving past stages of evolution... The inherent polysemy of natural signs is the indispensible instrument demanded by our mental complexity" (Fónagy 1979: 197).

Thus, it is not a question of why grammarians have resorted to phonetic metaphor. All human thinking and language uses it in the same way and for essentially the same ends--except that the scientist must then take the process one stage further by eventually discarding the metaphor and replacing it with literal, quantifiable observation. "The poet deliberately ignores a word; in that way, he suppresses a concept which already exists in order to express a new, individual vision of phenomena. The scientist... is searching for an object, a concept which only exists in the form of a gap, that is to say, a noted absence. At that moment, he takes the risk of a momentary regression in order to fix at least the imaginary, distorted mental reflection of the object which he

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is trying to grasp. He is in a hurry to unmotivate, to demystify the term he has just assigned to the unknown object. He insists on the arbitrary character of the term." (Fónagy 1979: 197).

This contrast can be seen in the metaphorical use of grammatical structures in poetry and the absence of such use in scientific language. The poet deliberately and expressively breaks the rules and thereby condenses several messages into one utterance. Scientists try to avoid this, strictly maintaining logical and syntactic structures: it is through their rigor and precision that they are able to express new ideas. Even though using natural language, they "try to stay at the highest evolutionary level of semantic structure," to distance themselves "from mythical implications and as far as possible reduce their polyvalence." Fónagy calls this "using language in a progressive, rather than regressive, manner" (Fónagy 1979: 197—198).

In his later book, *La vive voix* (1983) Fónagy focuses on the performance aspect of speech; the last sections attempt to describe a science of performance (spoken poetry, drama) and though fascinating in themselves are not relevant to this survey. Parts of his earlier book are also found much summarized in the later one. However, what is both new and relevant here is his theory of phonetic symbolism, which appears to have developed directly from Sapir's (1927) concepts of the expressive and referential dimensions of language.[23] But whereas Sapir talks of the *intertwining* of the dimensions in speech, Fónagy's claim is that for every utterance (i.e.,, sequence of segments) a secondary, emotional message is *grafted* onto the string. Depending on the emotion expressed, this secondary, parasitical message will change—in a predictable way—

[23] Surprisingly, reference to Sapir's 1927 article appears in neither the bibliography of *La métaphore en phonétique* (1979), nor in that of *La vive voix* (1983).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the articulatory and acoustic natures of the host segments it is riding on. Further, he asserts that in speech, we never get a phoneme in its "pure form" (Fónagy 1983: 23), since in the moment that it is being realized by the vocal organs, another message (communicating the speaker's attitude towards utterance and context) inevitably overlays the first. Indeed, Fónagy claims that not just our speech but every act of ours is done with "a parasitic activity (hidden) grafted onto the activity undertaken with a specific end, conscious" (Fónagy 1983: 25).

He gives a number of rules for producing (phonetic) gestural signals, the most general of which are as follows: "1. Voluntary reproduction of the vocal symptoms of an emotion signals the presence of that emotion... 2. Speech organs may represent, symbolize other animate or inanimate objects which are associated with them by similarity or by functional analogy... 3. The principle of isomorphism of expression and of content demands that to different degrees of semantic intensity there correspond different degrees of intensity on the plane of sound expression..." (Fónagy 1983: 17-18).

Fónagy observes that emotions show their influence on speech sounds in highly predictable ways; his implication is that these tendencies are universal. For example, anger produces jerky tongue movements, very fast transitions and a tendency to stay longer in target positions, tending to extremes: thus the high vowels become very closed and the low vowels very open. All angrily enunciated consonants display more pressure and contact due to greater than normal muscular tension. Additionally, in anger the tongue is more retracted for both vowels and consonants "which lends a darker tone to sounds" (Fónagy 1983: 31). Hatred further increases muscular tension and the consequent area of contact for consonants. Tenderness, on the other hand,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

produces a more relaxed articulation with slower transitions and shorter steady-state positions; so that all movements tend to be more uniform in length, with more flow and less jerkiness. Tender articulation is relatively advanced, "which lends overall, a brighter timbre to sounds" (Fónagy 1983: 32). [24]

Why are the emotions phonetically embodied in just this way? At this point, Fónagy declares himself a staunch Darwinian: "The theory proposed by Darwin seems to me invariably the best framework for an interpretation of vocal expression of emotive attitudes: the world surrounding primitive man and his ancestors... By putting the emotions back into their natural milieu, Darwin transforms them into activities directed towards precise goals... complex activities, useful in certain circumstances *and which are automatically triggered by all analogous situations*" (Fónagy 1983: 32-33, emphasis added). Thus the manifestations of anger and hatred correspond to preparations for fighting, hatred being more interiorized and no longer comporting the direct threat of biting and swallowing the victim. This demonstrates the operation of the Darwin's (1955) first priniciple, that of serviceable associated habits; the opposite emotion of tenderness is then phonetically embodied according to Darwin's second principle, that of antithesis.

As Sapir does, Fónagy accords a great deal of expressivity to glottal activity and devotes much discussion in his book to the emotional use of F_0, but in terms of vowel symbolism, the essential material is in his very long third chapter, "The pulsional bases of phonation." He remarks here

[24] The extent to which emotions can change speech performance is Ilustrated in a series of drawings taken from an experiment using the same 6 phrases uttered in a variety of emotions and recorded on "film radiologique." Fonagy demonstrates how under the influence of certain feelings, /i/ tongue position can lower to /e/ while other emotional states will raise /e/ to the position for /i/ (Fonagy 1983: 35-42).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that vowels are often "assigned a sex... velar vowels... (are considered as) male... anterior vowels... as female... The convention assigning a metaphorical term to a specific vocal quality is not necessarily arbitrary... It is very unlikely that Greek, Mongolian and Japanese grammars use analogous metaphors for similar phonetic traits without their being a natural relation between the acoustic phenomenon and the figurative term..." (Fónagy 1983: 52).

Fónagy points out the association between back vowels and sadness due to acoustic (i.e., low frequency) elements, giving a quality of "somberness," but he reminds the reader that there are also articulatory reasons in the "relatively retracted and depressed position of the tongue" (Fónagy 1983: 73). On the other hand, tongue position for /i/ is more advanced and higher, corresponding to the articulatory setting for joyful speech. These points are illustrated in the drawings of the articulation of different emotionally-charged utterances, taken from a performance experiment in which the same six phrases are enunciated by an actor depicting such emotions as tenderness, disdain, hatred and so on. X-ray photographs showed marked articulatory changes in the same segment, depending on the emotion being communicated.

While admitting that the causes of some phonically-engendered metaphors are not as simple to locate or explain as the preceding examples of the back vowels and /i/, Fónagy once again rejects arbitrariness as an explanation. Perhaps, he speculates, the sound has triggered associations that pass through the unconscious before ending as a metaphor: "Phonetic analysis operates... on a preconscious level, and (we know that) preconscious mental operations undergo the attraction of the unconscious" (Fónagy 1983: 75). However, his project is to trace "the associations aroused

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

by the phonatory act... in the hopes of disengaging the unconscious background of phonation" (Fónagy 1983: 78). There follows a discussion of sweetness and oral eroticism, and their association with palatalizaion and anteriority. He reiterates the connection between /i/ and sweetness (his reasons are that it is in the palatalizing position and makes use of the portion of the tongue where sweet taste is processed); /u/ is associated with bitterness since it uses the back of the tongue where bitter taste is registered.

The connection between the primitive (both ontogenetic and phylogenetic) and the metaphorical is clearly made: poetry is "directed regression," as anger or other emotions are regressions, and in it "the existence and importance of vocal mimicry become evident," for example "'soft' 'sweet' consonants dominate in poems of tenderness... This remotivation is connected to a structural and functional regression which brings verbal communication back to the stage of gestural, prelinguistic communication or even back to the precommunicative, ludic stage which in ontogenesis corresponds to babbling. The term *expressive*, so often used and little explained, encompasses an evolution that stretches over several tens of thousands of years... The expressive signal represents an intermediate stage: it belongs to the most evolved system of communication, to language, but also takes up the archaic, autistic stage of *acting* in which physical movements serve to reduce tension directly, immediately" (Fónagy 1983: 148).

Thus, "One denies or accepts the distant past, inherent to the present, the existence of an archaic communication integrated into the most evolved communication, *the persistence of verbal childishness behind the facade of an adult mentality*" (Fónagy 1983: 202, emphasis added).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

There is a further connection made between this expressive dimension of language and "a magical, infantile conception of the universe..."[25] in which "the separation of the self and the exterior world is still very imperfect: we are far from a clear distinction between the objective and the subjective which characterized thought and language properly speaking... Thus one chokes (i.e., speaks in a choked voice) in order to choke the adversary" (Fónagy 1983: 204-205).

In its wonderfully integrated quality, its welding together of psychology and poetry, stylistics and phonetics, Fónagy's work offers the most coherent explanation that I have yet found of the nature of phonetic symbolism and of the reasons for its existence. If he over-stresses its universality and says too little about the possible phonosymbolic influences of specific languages or cultural factors, his theory will still allow for these influences. Most important, he, like Sapir, emphasizes the essentially primitive nature of phonetic metaphor and encourages us to acknowledge and explore it. And finally, he gives a most reasonable explanation for the heat of the debate over phonetic symbolism, for the resistance shown towards the whole concept and for the way that its opponents have overrelied on the 'rational' principle of arbitrariness. "A shiver of mystery signals that we are on the threshold of the unconscious... We deny or accept the distant past underlying the present, the existence of an archaic communication integrated into the most evolved communication, the persistence of verbal childishness behind the façade of an adult mentality. We forget too easily that language, that intellectual performance placed deservedly so high—'the supreme gift of men' said Goethe—is at the same time a charming play of organs, a domesticated cheeping. The foregrounding of the 'nocturnal side of language,' as it does that of

[25] This connection is also made by Werner and Kaplan (1963).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

mentality in general, acts painfully by bruising our vanity. The millenary struggle between the principles of *physei* and *thesei* probably dramatizes this ambivalence" (Fónagy 1978: 159).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

Akamatsu, T. 1997 *Japanese phonetics: theory and practice* (Lincom studies in Asian linguistics 3) Munchen, Newcastle: Lincom Europa

Alspach 1917 On the psychological response to unknown proper names, *American Journal of Psychology* 28: 436—443

Asher, J. J. 1964 Vision and audition in language learning, *Perceptual and Motor Skills* 19: 255--300

Asher, J. J. 1965 Comment: phonetic symbolism re-examined, *Perceptual and Motor Skills* 20: 824

Atzet, J. and Gerard, H.B 1965 A study of phonetic symbolism among native Navajo speakers. *Journal of Personality and Social Psychology* 1: 524—528

Baindurašvili, A. 1967 Results of some experimental studies on naming, *International Journal of Psychology* 2 (3): 161—169

Barik, H.C. 1964 Some critical comments on visual presentation in word-matching studies of phonetic symbolism, *Language and Speech* 12: 175—179

Barrutía, R and A. Schweigler 1994 *Fonética y fonología españolas: teoría y práctica* New York: John Wiley & Sons

Bentley, M. and E.J. Varon 1933 An accessory study in phonetic symbolism, *American Journal of Psychaiatry* 45: 76—86

Berlin, B. 1997 Tapir and Squirrel: further nomenclature meanderings towards a universal sound-symbolic bestiary (invited lecture presented before the 20th annual meeting of the Society of Ethnobiology, University of Georgia, 26-29 March 1997)

Birch, D. and M. Erikson 1958 Phonetic symbolism with respect to three dimensions from the semantic differential, *Journal of General Psychology* 58: 291—297

Bloomfield, L. 1933 *Language*, New York: Henry Holt & Company

Bloomfield, M. 1895 On assimilation and adaptation in congeneric classes of words, *American Journal of Philology* 16, 409—434

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bolinger, D.L. 1965 *Forms of English: accent, morpheme, order*, Cambridge, MA: Harvard University Press

_____ 1992 Sound symbolism, *International Encyclopedia of Linguistics*, 1st ed., New York: Oxford University Press

Borden, G.J., K. Harris and L. Raphael 1994 *Speech Science Primer* Baltimore: Williams and Wilkins

Brackbill, Y. and Little, K. B. 1957 Factors Determining the Guessing of Meanings of Foreign Words, *Journal of Abnormal and Social Psychology* 54: 312—318

Brosnahan, L.F and B. Malmberg 1970 *Introduction to Phonetics*, Cambridge: Cambridge University Press

Brown, G. 1994 Sounds and meanings, *IATEFL Annual Conference Report*

Brown, R.W., Black, A.H. and Horowitz, A.E. 1955 Phonetic symbolism in natural languages, *Journal of Abnormal and Social Psychology* 50: 388—393

_____ 1958a *Words and things*, New York: The Free Press, Macmillan & Company

_____ 1958b Is a boulder sweet or sour? *Contemporary Psychology* 3: 113--115

_____ and Nuttall, R 1959 Methods in phonetic symbolism experiments, *Journal of Abnormal and Social Psychology* 59: 441--445

Carroll, J.B. 1959 Review of 'The measurement of meaning,' *Language* 35: 58-77

Castillo, C., O.F. Bond and B. M. García 1966 *The University of Chicago Spanish-English English-Spanish dictionary*, New York: Washington Square Press

Chastaing, M. 1958 Le symbolisme des voyelles, signification de i, *Journal de Psychologie* 55: 403—423 & 461—481

_____ 1962 La brillance des voyelles, *Archivum Linguisticum* 14: 1—13

Chatman, S. 1957 Linguistics, poetics and interpretation: the phonemic dimension, *Quarterly Journal of Speech* 63 (3) 248-25

Chiba, T. and Kajiyama, M. 1958 *The vowel, its nature and structure* Tokyo: Phonetic Society of Japan

Cho, S. B. 1967 *A phonological study of Korean*, Uppsala: Almqvist and Wiskells

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chomsky, N. and M. Halle, 1968 *The sound pattern of English*, New York: Harper & Row

Chung, H.Y, S. Makino and K. Kido 1988a Analysis and recognition of Korean isolated vowels using formant frequency, *The Journal of the Acoustical Society of Japan* Vol. 9 #5, 225—232

_____ 1998b On the listening tests using the natural speech of eight vowels, *The Journal of the Acoustical Society of Japan* Vol. 9 # 6, 305—308

Ciccotosto, N. 1991 *Sound symbolism in natural languages*, unppublished Ph.D. dissertation, University of Florida, Gainesville, Florida

Claparède, E. 1914 De la répresentation des personnes inconnues, et des lapsus linguae, *Archives de Psychologie* 14: 301—304

Clarke, E. C. and S. S. Yun 1950 *Introduction to spoken Korean* New Haven, CT: Institute of Far Eastern Languages, Yale University

Colton, M.A. 1909 *La phonétique castellane: traité de phonétique déscriptive et comparative*, Paris: Lievens à St. Maur

Crockett, D. 1970 Secondary onomatopoeia in Russian, *Word* 26: 107—113

Croft, L.B. 1978 The mnemonic use of linguistic iconicity in teaching language and literature, *Slavic and East European Journal* 22 (4): 509—518

Davis, R. 1961 The fitness of names to drawings: a cross-cultural study in Tanganyika, *British Journal of Psychology* 52: 259—268

Darwin Charles E. 1955 *The expression of the emotions in man and the animals*, New York: Philosophical Society

Del Carmen McNair, M. 1996 *An acoustical analyis of Panamanian vowels*, unpublished master's thesis, University of South Florida

De Reuse, W.J. 1986 The lexicalization of sound symbolism in Santiago del Estero Quechua, *International Journal of American Linguistics* 52: 54—64

De Vito, J. and Civikly, J.M. 1972 Some semantics of repetition: an experiment in phonetic symbolism, *Journal of Communication* 22: 39—47

Diffloth, G. 1994 i: big, a: small, in Hinton, L., Nichols, J. and Ohala, J. J., eds *Sound symbolism*, Cambridge: Cambridge University Press

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Eberhardt, M. 1940 A study of phonetic symbolism of deaf children, *Psychology Monographs* 52: 23—41

English, G. 1916 On the psychological response to unknown proper names, *American Journal of Psychology* 27: 430—434

Ervin-Tripp, S.M. and D.I Slobin, 1966 Psycholinguistics, *Annual Review of Psychology* 17: 454—456

Escardt, T. and W. C. Oh 1995 *Hippocrene handy dictionary of Korean* (2nd edition) New York: Hippocrene Books

Fischer-Jørgensen, E. 1978 On the universal character of phonetic symbolism with special reference to vowels, *Studia Linguistica* 32: 80—90

Florez, L. 1951 *La pronunciación del español en Bogotá*, Bogotá: Publicaciones del Instituto Caro y Cuervo 8

Fónagy, I. 1979 *La métaphore en phonétique*, Ottawa: Marcel Didier

_____ 1983 *La vive voix*, Paris: Payot

Fox, A. 1982 Sound symbolism in Semitic, *Cahiers de l'Institut de Linguistique de Louvain* 8: 55—61

Fox, C.W. 1935 An experimental study of naming, *American Journal of Psychology* 47: 545—579

French, P. 1977 Towards an explanation of phonetic symbolism, *Word* 28: 305—322

Fry, D. B. 1979 *The physics of speech* Cambridge: Cambridge University Press

Fudge, E. 1970 Phonological structure and "expressiveness," *Journal of Linguistics* 6: 161—315

Gamkrelidze, T.V. 1974 The problem of "l'arbitraire du signe" *Language* 50: 102—110

Garver, L.N, Gleason, J.M. and Washburn M.F. 1915 The source of affective reactions to articulate sounds, *American Journal of Psychology* 26: 292—295

Gebels, G. 1969 An investigation of phonetic symbolism in different cultures, *Journal of Verbal Learning* 8: 310—324

Gimson, A.C. 1989 *Introduction to the pronunciation of English*, (4th edition) London: Edwin Arnold

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Grammont, M. 1946 *Traité de phonétique*, Paris: Delagrave

Guirao, M and A. M. Borzone de Manrique 1975 Identification of Argentine Spanish vowels, *Journal of Psycholinguistic Research*, Vol. 4 #1 17—25

Guest, E. 1838 *A history of English rhythms*, London: William Pickering

Haiman, J. 1983 Iconic and economic motivation, *Language* 59: 781—819

Halberstam, B. 1998 *Vowel normalization: the role of F_0 and the upper formants*, Unpublished dissertation, CUNY Graduate Center, New York City

Hall, K.R.L. and Oldfield, R.C. 1950 An Experimental Study on the Fitness of Signs to Words, *Quarterly Journal of Experimental Psychology* 2: 60—70

_____ 1957 The fitness of signs to words, *British Journal of Psychology* 42: 21—33

Hamano, S.S. 1986 *The sound symbolic system of Japanese*, unpublished Ph.D dissertation, University of Florida, Gainesville, Florida

Han, M. S. 1961 *Japanese phonology: an analysis based upon sound spectrograms*, unpublished dissertation, University of Texas

Heise, D. R. 1966 Sound-meaning correlations among 1,000 English words, *Language and Speech* 9: 14—27

Hevner, K. 1937 An experimental study of the affective value of sounds in poetry, *American Journal of Psychology* 49: 419—434

Hinton, L., Nichols, J. and Ohala, J. J., eds 1994 *Sound symbolism*, Cambridge: Cambridge University Press

Holland, M.K. and Wertheimer, M. 1964 Some physiognomic aspects of naming, or Maluma and Takete revisited, *Perceptual and Motor Skills* 19: 111—117

Homma, Y. 1973 An acoustic study of Japanese vowels, *Onsei no Kenkyu/Study of Sounds* 16: 347—368

Hornbostel, E. M. von 1927 The unity of the senses, *Psyche* 7: 83—89

Householder, F.W. 1946 On the problem of sound meaning: an English phonestheme, *Word* 2: 83—84

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

_____ 1960 From the viewpoint of linguistics: opening statement, in Sebeok, T. ed, *Style in language* Cambridge, MA, M.I.T. Press: 339—349

Hunt, H.T. 1985a Cognition and states of consciousness: the necessity for empirical study of ordinary and nonordinary consciousness for contemporary cognitive psychology, *Perceptual and Motor Skills* 60: 239—282

_____ 1985b Relations between the phenomena of religious mysticism (altered states of consciousness) and the psychology of thought: a cognitive psychology of states of consciousness and the necessity of subjective states for cognitive theory, *Perceptual and Motor Skills* 61: 911—961

Huysmans, J-K. 1959 *Against nature*, London: Penguin Books

Hyman, L. M. 1957 *Phonology: theory and analysis*, New York: Holt, Rinehart & Winston

Hymes, D.H. 1960 Phonological aspects of style: some English sonnets, in Sebeok, T. ed, *Style in Language* Cambridge, MA, M.I.T. Press: 109—131

Iritani, T. 1969 Dimensions of phonetic symbolism: an inquiry into the dynamic expressiveness of features in the symbolization of nonlinguistic sounds, *International Journal of Psychology* 4 (1): 9—19

Irwin, F.W. and Newland, E. 1940 A genetic study of the naming of visual figures, *Journal of Psychology* 9: 3—16

Isshiki, M. 1965 A contrastive analysis of English and Japanese vowel phonemes *Onsei no Kenkyu/Study of Sounds* 8: 193—211

Jakobson, R. Fant, G. and Halle, M. 1951 *Preliminaries to speech analysis: the distinctive features and their correlates*, Cambridge, MA: MIT Press

_____ 1960 Closing statement: linguistics and poetics in Sebeok, T. ed, *Style in language* Cambridge, MA, M.I.T. Press

_____ 1968 Agreement between the systems of sound and colour, *Child Language, Aphasia & Phonological Universals, Janua Linguarum* 72, The Hague, Mouton 82—84

_____ and Halle, M. 1971 *Fundamentals of language*, The Hague: Mouton

_____ 1978 *Language in literature*, Cambridge MA: Belknap Press of Harvard University

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Jespersen, O. _____ 1922 *Language: Its nature, development, origin,* New York, London: George Allen & Unwin, p 397—411

_____ 1933 The symbolic value of the vowel i, in Linguistica: *Selected writings of Otto Jespersen,* Copenhagen: Levin and Munksgard 283-303 London: George Allen & Unwin

Johnson, R.C. 1967 Magnitude symbolism of English words, *Journal of Verbal Learning and Verbal Behavior* 6: 508—511

_____ Suzuki, N.S. and Olds, W.K. 1964 Phonetic symbolism in an artifical language, *Journal of Abnormal and Social Psychology* 69: 233—236

Jones, D. 1978 *The pronunciation of English,* Cambridge: Cambridge University Press

Joos, M. 1946 Acoustic phonetics (supplement to *Language*), *Language monograph 23*

Karwoski, T.F, H. S. Osbert and C.E Osgood 1942 Studies in synesthetic thinking II. The role of form in visual response to music, 26: 199—222 *The Journal of General Psychology*

Kang, S.K. 1966 Acoustic analysis of South & North Korean vowels, *Ohak Yonku/Language research* 32: 1—18 (Korean, but with abstract in English)

Katz, A.N. 1986 Meaning conveyed by vowels: some reanalyses of word norm data, *Bulletin of the Psychonomic Society* 24: 15—17

Klank, L.J.K., Huang, Y. and Johnson, R.C. 1971 Determinants of success in matching word pairs in tests of phonetic symbolism, *Journal of Verbal Learning* 10: 140—148

Knapp, R.H. and Ehlinger, H.J. 1968 Study of semantic associations to the alphabet, *Perceptual and Motor Skills* 26: 140—148

Köhler, M. 1947 *Gestalt psychology,* New York: Liveright

Kollarits 1914 Observation de psychologie quotidienne, *Archives de Psychologie* 14: 225—247

Kunihara, S. 1971 Effects of the expressive voice on phonetic symbolism, *Journal of Verbal Learning and Verbal Behavior* 10: 427—429

Kurath, H. and R. I. McDavid 1961 *The pronunciation of English in the Atlantic states: based upon the collections of the Linguistic Atlas of the eastern United States,* Ann Arbor: University of Michigan Press

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Ladefoged, P. 1962 *Elements of acoustic phonetics*, Chicago & London: University of Chicago Press

_____ 1982 *A course in phonetics*, second edition, New York: Harcourt, Brace, Jovanovich Inc.

Langer, J. and Rosenberg, B.G. 1964 Non-verbal representation of verbal referents, *Perceptual and Motor Skills* 19: 363—370

Lapolla, R. 1994 An experimental investigation into phonetic symbolism as it relates to Mandarin Chinese, in Hinton, L., Nichols, J. and Ohala, J. J., eds *Sound symbolism*, Cambridge: Cambridge University Press

Langer, S.K. 1942 *Philosophy in a New Key*, Cambridge, MA: Harvard University Press

Laver, J. 1994 *Principles of phonetics*, Cambridge: Cambridge University Press

Lee, C. H. 1955 *Practical Korean grammar*, Seattle: University of Washington Press

Lee, D.Y. 1981 *An introduction to East Asian and Tibetan linguistics and culture*, Bloomington, IN: Eastern Press

Lewis, M. 1986 *The English verb: an exploration of structure and meaning*, Hove: Language Teaching Publications

Lukoff, F. 1945 *Spoken Korean*, Linguistic Society of America: Henry Holt & Company

Lynch, J. J. 1953 The tonality of lyric poetry: an experiment in method, *Word* 9: 211—214

Macdermott, M.M. 1940 *Vowel sounds in poetry: their music and tone-color*, London: Kegan Paul

Mackay, I. 1987 *Phonetics: the science of speech production*, second edition, Austin, Texas: Pro-ed

Magnus, M. 1999 *Gods of the word*, Kirksville, MO: Thomas Jefferson University Press

Mandelker, A. 1982 *New research in phonetic symbolism: the poetic context*, unpublished doctoral dissertation, Department of Slavic Languages and Literatures, Brown University

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Maltzmann, I., Morrisett, L.J. and Brooks, L. 1956 Critique and notes: an investigation of phonetic symbolism, *Journal of Abnormal and Social Psychology* 53: 245—251

Marchand, H. 1959 Phonetic symbolism in English word-formation, *Indogermanische Forschungen* 64: 146-168, 256—277

_____ 1969 *English word formation*, Munich: C. H. Beck'sche Verlagsbuchhandling

Markel, N.N. and Hamp, E.P. 1961 or 62 Connotative meanings of certain phoneme sequences, *Studies in Linguistics* 15: 47—61

Marks, L.E. 1978 *The unity of the senses: interrelations among the modalities*, New York: Academic Press

_____1982a Synesthetic perception and poetic metaphor, *Journal of Experimental Psychology: Human Perception and Performance* 8: 15—23

_____ 1982b Bright sneezes and dark coughs, loud sunlight and soft moonlight, *Journal of Experimental Psychology: Human Perception and Performance* 8: 177—193

_____ 1987 On cross-modal similarity: auditory-visual interactions in speeded discrimination, *Journal of Experimental Psychology: Human Perception and Performance* 13: 384—394

Martin, S.E. 1962 Phonetic symbolism in Korean. In *Uralic and Altaic series* edited by N. Poppe, Bloomington, Indiana: Indiana University Press: 177—189

_____ 1992 *A reference grammar of Korean: a complete guide to the grammar*, Rutland, VT: C. E. Tuttle

Masson, D. I. 1952 Synesthesia and sound spectra, *Word* 8: 39—41

Master-Mcneil 1996 *Sound symbolism and naming*, White paper available from the Master-Mcneil Advertising Company on the Internet

Mateo, A. 1988 Experimento sobre el tono intrinseco de las vocales castellanas, in *Estudios de Fonética experimental* 3, Barcelona: Laboratorio de Fonética, Facultad de Filología, Universidad de Barcelona

Mendez, A. 1982 Production of American English and Spanish vowels, *Language and Speech*, Vol. 2 part 2

McMurry, G.A. 1958 A study of "fittingness" of signs to words by means of the semantic differential, *Journal of Experimental Psychology* 56: 310—314

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

_____ 1960 Meaning associated with the phonetic structure of unfamiliar foreign words, *Canadian Journal of Psychology* 14: 166—174

Mendez, A. 1982 Production of American English and Spanish vowels, *Language and speech* 25 II : 191—197

Miron, M.S. 1961 Cross-linguistic investigations of phonetic symbolism, *Journal of Abnormal and Social Psychology* 62: 623—630

Morimoto, Y. 1988 The five vowels of Castillian and Japanese: a spectrographic analysis, *Sophia Linguistica* 23/24 181—191

Morton, E. 1994 Sound symbolism and its role in non-human vertebrate communication in Hinton, L., Nichols, J. and Ohala, J. J., eds *Sound symbolism*, Cambridge: Cambridge University Press

Murakami, Y. 1982 The facilitation effects of phonetic symbolism on paired associate learning, *Japanese Psychological Research* 24: 62—67

Murdock, G.P. 1959 Cross-language parallells in parental kin terms, *Anthropological Linguistics* 1: 1—6

Navarro Tomás, T. 1946 *Estudios de fonología española*, New York: Las Americas Publishing Company

_____ 1948 *El español en Puerto Rico: contribución a la geographia linguística hispanoamericana*, Río Pedras: Universidad de Puerto Rico

_____ 1967 *Manual de pronunciación española*, Madrid: Publicaciones de la Revista de Filología Española, III

Newman, S.S. 1933 Further experiments in phonetic symbolism, *American Journal of Psychology* 45: 53—75

Nichols, J. 1971 Diminutive consonant symbolism in Western North America, *Language* 47: 826—848

O'Boyle, M. W., D.A. Miller and F. Rahmani, 1987 Sound-meaning relationships in speakers of Urdu and English: evidence for a cross-cultural phonetic symbolism, *Journal of Psycholinguistic Research* 16: 273—288

Odbert, H.S., T.F. Karwoski and A.B. Eckerson 1942 Studies in synesthetic thinking I: musical and verbal associations of color and mood *The Journal of General Psychology* 26: 153—173

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Ohala, J 1994 The frequency code underlies the sound-symbolic use of voice pitch, in Hinton, L., Nichols, J. and Ohala, J. J., eds *Sound symbolism*, Cambridge: Cambridge University Press

Onishi, M. 1981 *A grand dictionary of phonetics*, The Phonetic Society of Japan

Orr, J. 1944 On some sound values in English, *British Journal of Psychology* 35: 1—8

Osgood , C.E., and G.J. Suci. 1955 Factor analysis of meaning, *Journal of Experimental Psychology* 50: 325—338

_____, G.J. Suci and P.H. Tannenbaum, 1957 *The measurement of meaning*, Urbana: University of Illinois Press

Oyama, T. and J. Haga 1963 Common factors between visual and phonetic symbolism, *Psychologia* 6: 131—144

Paget, R. 1927 *The origin of language*, Psyche 7: 35—39

_____ 1930 *Human speech*, New York: Harcourt Brace

Pecjak, V. 1970 Verbal synesthesiae of colors, emotions and days of the week, *Journal of Verbal Learning and Verbal Behavior* 9: 633—626

Peterson, G.E. and Barney, H.L. 1952 Control methods used in a study of the identification of vowels, *Journal of the Acoustical Society of America* 24 #2: 175—184

_____ and Lehiste, I. 1960 Duration of syllabic nucleii in English, *Journal of the Acoustical Society of America* 32 693—703

Plato 1892 Cratylus, *The dialogues of Plato*, Vol.1, Oxford: Clarendon Press

Piaget, J. 1955 *The language and thought of the child*, Cleveland: Meridian Books

_____ 1979 *The child's conception of the world* Totowa, NJ: Littlefield, Adams & Co.

Pickett, J. M. 1980 *The sounds of speech communication* Baltimore: University Park Press

Priestly, T. 1994 On levels of analysis of sound symbolism in poetry, with an application to Russian poetry in Hinton, L., Nichols, J. and Ohala, J. J., eds *Sound symbolism*, Cambridge: Cambridge University Press

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Quenk, A.T. 1964 *Phonic shift as a function of inappropriate stimulus conditions*, unpublished Ph.D. dissertation, Berkley: University of California

Quilis, A 1981 *Fonética acústica de la lengua española*, Madrid: Editorial Gredos

Ramstedt, G. J. 1939 A Korean grammar, *Mémoires de la societé finno-ougrienne* LXXXII Helsinki

Raphael, L., Bell-Berti, F., Collier, R., and Baer, T. 1979 Tongue position in rounded and unrounded front vowel pairs, *Language and Speech* 22: 37—48

Reichard, G., Jakobson, R. and Werth, E. 1949 Language and synesthesia, *Word* 5: 224—234

Roblee, L. and Washburn, M.F. 1912 The affective values of articulate sounds, *American Journal of Psychology* 23: 579—583

Romero, J. 1988 Campos de dispersión auditivos de las vocales del castellano. Percepción de las vocales, in *Estudios de fonética experimental* 3, Barcelona: Laboratorio de Fonética, Facultad de Filología, Universidad de Barcelona

Rosenblum and Pinker 1983 Word-magic revisited, *Child Development* 54: 773—780

Rudmin, F. and Cappelli, M. 1983 Tone-taste synesthesia: a replication, *Perceptual and Motor Skills* 59: 118

Ryalis, J.H. 1986 Synaesthesia: a principle for the relationship between the primary colors and the cardinal vowels, Semiotica 58, 1—2 : 10—121

Saka, P 1988 *Sound symbolism and the lexicon*, unpublished paper in the dissertation author's posession

Sapir, E. 1927 Language as a form of human behavior, *The English Journal* 16: 421—433

_____ 1929 A study in phonetic symbolism, *Journal of Experimental Psychology* 12: 225—239

Saussure, F. de 1915, 1959 *Course in general linguistics*, New York: McGraw-Hill

Schnitzer, M. L. 1997 *Fonologia contrastiva español-inglés*, San Juan, PR: Piedras Press

Serralta, F. and V. Tusón 1967 *Phonétique espagnole pratique*, Université de Toulouse: Institut d'études hispaniques

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sherman, L.A. 1893 *Analytics of literature* Boston: Ginn

Shtern, A. 1967 An objective study of the subjective evaluation of speech sounds, *Soviet Psychology IASP* 7 (3): 17—18

Siegel, A.W., Silverman, I. and Markel N.N. 1967 *On the effect of mode of presentation on phonetic symbolism*, Journal of Verbal Learning and Verbal Behavior 6: 171—173

Skelton, R. 1950 *A spectrographic analysis of Spanish vowel sounds*, unpublished dissertation, University of Michigan

Slobin, D.I. 1968 Antonymic phonetic symbolism in three natural languages. *Journal of Personality and Social Psychology* 10: 301—314

Solomon, L.N. 1958 Semantic approach to the perception of complex sounds, *The Journal of the Acoustical Society of America* 30: 421—425

_____ 1959a Search for physical correlates to psychological dimensions of sounds, *The Journal of the Acoustical Society of America* 31: 492—497

_____ 1959b Semantic reactions to systematically varied sounds, *The Journal of the Acoustical Society of America* 31: 986—990

Sorzano Jorrín, L. 1928 *Fonética elemental*, Habana: Cultural S.A.

Smith, L.P. 1912 *The English language* London: Williams & Norgate

Stevens, C. De Vito, J.A. and Isaacson, N. 1969 *Phonetic symbolism and audience perception*, Southern Speech Journal 34: 183—196

Strange, W., R. Akahane-Yamada, R Kubo, K Nishi, S. A. Trent and J. J. Jenkins 1998, *Perceptual assimilation of American English vowels by Japanese listeners*, Journal of Phonetics, Vol. 26 # 4: 311—344

Tanz, C. 1971 Sound symbolism in words relating to proximity and distance, *Language and Speech* 14: 266—276

Tarte, R. D. and Barritt, L.S. 1971 Phonetic symbolism in adult native speakers of English: three studies, *Language and Speech* 158—168

Tarte, R.D. 1974 Phonetic symbolism in adult native speakers of Czech, *Langauage and Speech* 17: 87—94

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

_____ and O'Boyle, M.W. 1982 Semantic judgements of compressed monosyllables: evidence for phonetic symbolism, *Journal of Psycholinguistic Research* 11: 183—196

_____ 1982 The relationship between monosyllables and pure tones: an investigation of phonetic symbolism, *Journal of Verbal Learning and Verbal Behavior* 21: 352—360

Tartter, V. C. 1980 Happy talk: perceptual and acoustic effects of smiling on speech, *Perception and psychophysics*, Vol. 27 # 1: 24—27

Taylor, I.K. 1963 Phonetic symbolism re-examined, *Psychological Bulletin* 60: 200—209

_____ 1965a Phonetic symbolism and vocabulary learning, *Perceptual and Motor Skills* 20: 842

_____ 1965b Note on phonetic symbolism, *Perceptual and Motor Skills* 20: 803—804

_____ 1967 An anatomy of words used in a word-matching phonetic symbolism experiment, *Journal of General Psychology* 76 (2): 231—239

_____ and Taylor, M.M. 1962 Phonetic symbolism in four unrelated languages, *Canadian Journal of Psychology* 16: 344—356

_____ and Taylor, M.M. 1965 Another look at phonetic symbolism, *Psychological Bulletin* 64: 413—427

Thorndike, E.L. 1945 On Orr's hypothesis concerning the front and back vowels, *British Journal of Psychology* 36: 10—14

Tolman, Albert H. 1887 The laws of tone color in the English language, *Andover Review* 7, 326—337

_____ 1904 The symbolic value of English sounds, in *Views about Hamlet*, New York: Houghton Mifflin

Tsuru, S. and Fries, H.S. 1933 A problem in meaning, *Journal of General Psychology* 8: 281—284

Vance, T. 1987 *An introduction to Japanese phonology*, Albany NY, State University of New York Press

Viglionese, P.C. 1985 The inner functioning of words: iconicity in poetic language, *Visible Language* 19: 373—386

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Waugh, L. 1993 Against arbitrariness: imitation and motivation reviewed with conseqences for textual meaning, *diacritics* Vol. 23 # 2: 71-87

Weiss, J.H. 1963a The role of meaningfulness versus meaning dimensions in guessing the meaning of foreign words, *Journal of Abnormal and Social Psychology* 66: 541—546

_____ 1963b Further study of the relation between the sound of a word and its meaning, *American Journal of Psychology* 76: 624—630

_____ 1964a Phonetic symbolism re-examined, *Psychological Bulletin* 61: 454—458

_____ 1964b The role of stimulus meaningfulness in the phonetic symbolism response, *Journal of General Psychology* 70: 255—263

_____ 1966 A study of the ability of English speakers to guess the meanings of non-antonymic foreign words, *Journal of General Psychology* 74: 97—106

Werner, H. & B. Kaplan 1963 *Symbol formation: an organismic-developmental approach to language and the expression of thought*, New York: Wiley

Wertheimer, M. 1958 The relation between the sound of a word and its meaning, *American Journal of Psychology* 71: 412—415

Wescott, R.W. 1971a Labiovelarity and derogation in English: a study in phonosemic correlation, *American Speech* 46: 123—137

_____ 1971b Linguistic iconism, Language 52: 461—478

_____ 1987 Holesthemes or phonesthemes twice over, *General Linguistics* (University Park, PA) 27: 67—72

Whalen, D. H. & A.G. Levitt 1995 The universality of intrinsic F_0 of vowels, *Journal of Phonetics* 23: 349—366

Williams. J.M. 1976 Synesthetic adjectives: a possible law of semantic change, *Language* 52: 461—478

Yang, B. 1996 A comparative study of American English and Korean vowels produced by male and female speakers, *Journal of Phonetics* Vol. 24 # 2: 245—261

Zamorra, V. 1960 *Dialectologia española*, Madrid: Editoriales Gredos

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.