REFERENCE IN THE DESCRIPTION OF FIGURES:

EFFECT OF CONTEXT ON LANGUAGE PRODUCTION

by

ESTHER MALCA RESNICK

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

1996

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

9/17/96
Date

Vivien C. Tartter
Chair of Examining Committee

9-20-96
Date

Kay Deaux
Executive Officer

Walter Ritter, Ph.D.

Jeffrey Rosen, PhD

Vivien Tartter, Ph.D.

Susan DeSanti, Ph.D.

Richard Gerrig. Ph.D
Supervisory Committee

The City University of New York

Abstract

REFERENCE IN THE DESCRIPTION OF FIGURES:

EFFECT OF CONTEXT ON LANGUAGE PRODUCTION

by

Esther Malca Resnick

Advisor: Professor Vivien Tartter

These studies explore the production of literal versus analogical (figurative) language through the description of geometric, abstract shapes. Subjects provided descriptions for later identification by two audience conditions: themselves and a peer. They were then asked to describe the figures in strictly literal (geometric) terms, in order to examine whether they were able to use an alternative description style to the one originally chosen. Subjects were asked to do this in both written and oral modalities, in order to study whether style/content varied across modality of production. Peers of this group were then asked to identify the figures using the descriptions generated by the language producers.

Results indicated that subjects used predominantly figurative language to describe the shapes for both audience conditions. In the "self" description condition, subjects used fewer words and proportionately more figurative language than they did in the "other" condition. They also used more specific more idiosyncratic nouns and less complex sentences. Subjects using these descriptions were able to identify figures produced under "self" and "other" conditions equally well, but were less accurate identifying the

descriptions generated under the imposed "literal" condition.

These results support the hypothesis that context plays a key role in language production in terms of influencing word choice and style of language. Also, it was shown that people tailor descriptions specifically for their perceived target audience, often providing more than was necessary for successful comprehension. Results are discussed within a constructionist framework.

ACKNOWLEDGMENTS

I would like to thank my committee, for their efforts individually and as a unit whose sum was even greater than the parts: Walter Ritter, whose questions and suggestions helped me formulate a dissertation that (hopefully) kept sight of the larger picture; Jeffrey Rosen, for forcing me to learn to think; Vicky Tartter, for enthusiasm in my topic, and for making me work in spite my self-obstructionism by reminding me that life always intrudes, but you can't let it hold you up; my outside readers, who although they came into the picture relatively late, nonetheless played decisive roles in the endgame: Susan DeSanti, for her coding and linguistic analysis expertise; and Richard Gerrig, whose insightful and challenging comments and questions especially in the discussion portions helped me reframe this dissertation into a cohesive, structured whole. I am also grateful to Valerie, Lana,Hilary, Chris and Tara, my fellow students and co-members of Jeff's Monday dissertation seminar.

Thanks also to a number of others for the roles they played in the development and completion of my work: Chris Beemer and Peter Vavagiakis, for their help with the stimulus graphics; Chaim Tarshish and Ora Ezrachi for help (actual and emotional) with statistics when I needed it most; Suzy Joffe for linguistics; Glenn Goldfinger and Robin Blank, Seema Morrell and Rachel Levine for enthusiastic recruiting; Leda Kydoniefs and Tamar K. Goldstein for the use of their homes and shoulders when my own weren't enough. I cannot mention by name all my friends who stoically stood by me, and offered support and encouragement, and somehow still remained my friends throughout. Thanks to all my co-workers at the NYU Aging & Dementia Research Center for providing the intellectual edge.

And special thanks to my family, my parents and siblings, for their inexplicable but persistent belief in my ability to complete this research. And to those relatives whose threats of homicide did their job.

In memory of Katja
who stood by (and sat on) my dissertation research.

TABLE OF CONTENTS

LIST OF TABLES

"When I say two different things, I mean two different things by them."

(MacWhinney, 1989, quoting Bollinger).

INTRODUCTION

OVERVIEW

In everyday speech, we produce two to three words per second, and there are occasional bursts of up to seven words per second (Bock & Levelt, 1994). To speak fluently, we must select and retrieve words from our mental lexicons virtually simultaneously with, if not prior to, the speech act. Each word must be selected, modified appropriately, placed in the proper syntactic framework, and pronounced. Thus the study of language production encompasses the mechanical processes of the speech mechanism, neuronal activity and stimulation of different areas of the brain, and, at a higher level, how these processes work in tandem with the speaker's conscious and unconscious intentions in his selection of words and phrases.

Language is manipulated by the speaker in different ways to represent concepts and emotions and thoughts, and to communicate them to others. While every word has its standard, accepted dictionary meanings, the same words in a different context can also represent new or alternative meanings. An almost infinite number of words can be used to refer to a particular object in the right circumstance, and conversely, while a particular word can assume different meanings, there is a fundamental principle, the Principle of Contrast, that no two word constitute perfect synonyms (e.g., Clark, 1992; MacWhinney, 1989). Each different situation elicits the most appropriate word or type of output, causing words to be interpreted in a specific manner appropriate to that context.

Sometimes the words or phrases may be literally untrue (e.g., it's raining cats and dogs), yet still are understood correctly by the listener. Furthermore, the same sentence in different contexts can be intended, and understood, completely, as in the sentence, "Elsie is a cow". The words or phrases chosen to recount a story or to describe a situation or picture can vary infinitely, depending on factors such as the relationship between the speaker and receiver, the current situation, and the intentions of the speaker, as well as individual differences, such as cultural background, level of education, and the intellectual abilities of the speaker.

Historically, research on meaning has focused primarily on its comprehension. Relatively little research effort has been concentrated upon aspects of production, such as *why* speakers select the words and phrases they do to best communicate their thoughts. Those studies that have studied language production have concentrated primarily on word frequency and grammatical content. Few studies, if any, have systematically looked at the specifics of word selection in relation to the context of its production.

Because language is so context-dependent and also so linked to the individual style of the speaker, it is difficult to study the conditions of its production systematically. One of the ways to minimize individual differences is to specify the context for the communicator (e.g., the target audience). Another is to provide the scenario for which the communication is about (e.g., the modality of communication and stimuli).

In the following pages, we will discuss various aspects of language production in the framework of figurative language and intention of communication. This dissertation will begin with a discussion of reference in communication, and how context and

common ground between participants in a dialogue influence how reference is established, with the best words selected for each situation. We will continue with a brief discussion about the complexity of word selection, including stages at which difficulties might occur, and including the intellectual functions of naming and vocabulary size. Then, using this framework of context determining production, we will briefly describe a traditional linguistic division, that of literal and nonliteral (figurative) language, and the vague boundary between them, if such a boundary exists at all. Several functions and purposes of what is commonly referred to as figurative language are suggested from the literature. We will then examine several external factors that might contribute to differences in production between individuals, such as age and sex, and modality of production, i.e., whether the communication was written or oral. Finally, we will describe the experiments performed herein, in which we explore the language of description of figures by an individual to either himself, or for an imagined audience. These studies examine some of the conditions that constrain and/or extend word meanings in hopes of further elucidating how language is linked to the conceptualization of experience.

REFERENCE, CONTEXT, AND COMMON GROUND

The fundamental purpose of language is effective transmission of information between communicating parties. To do so, a speaker needs to select and produce a string of words in such a way as to lead the listener to think that he means by that sentence the specific things he hopes to mean (Oden, et al., 1991). Obviously, there is no single way to communicate. Different situations, or contexts, require language particularly suited for

the need of the moment. The term "context" is a vague one that has been defined in various ways ranging from its denotative dictionary meaning to the more wide-ranging and nebulous, and serving multiple and varying purposes in different fields of study. For the purpose of studying language production, for the current study we will adopt the precise yet inclusive definition presented by Clark (1992), that context is information in an individual's possession, relevant to an occasion-specific particular process the person is engaged in, that must be able to interact with the process at hand. In sum, context is information that is available to a particular person for interaction with a particular process on a particular occasion (p65).

The first step in the process of description is to create reference, a means of referring to that which is being described. Effective communication between individuals requires an understanding of the contextual surroundings. Some features that help create context are dynamic, and vary with the situation, such as the relationship between the speaker and receiver and the intentions of the speaker, while others are more static, intrinsic influences, including individual differences such as cultural background, level of education, and vocabulary of the participants. Differences in these contextual features must be incorporated by the speaker, and may lead to modified styles of communication, more extended or more terse, more formal or more colloquial. Some of these factors will be elaborated upon herein, and will be included later as independent variables (or controlled for to minimize variability) in our experiment.

In the traditional view of reference and communication, only the role of the speaker was considered, independently of context; the respondent was considered

passive, as someone whose responsibility was to comprehend the intentions conveyed by the speaker. Later, the idea of active cooperation between the parties became emphasized as essential in the principles of conversation (e.g., Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1991; Martin, 1987), and this has become a major focus of current research. A conversational model which has evolved from the rules of both philosophers and linguists and the observations by sociologists focuses upon the coordination of the parties involved. Both speaker and the listener are acknowledged to play active roles in the communication of the reference; beyond the assumptions of truthfulness, relevance, and appropriateness (e.g., Clark, 1992), feedback from the recipient of the information serves to define how the speaker will proceed- whether to amplify his description, whether to try another means of communication, or whether the point has been understood, and it is unnecessary even to finish the thought (Clark & Wilkes-Gibbs, 1986; McCann & Higgins, 1992). In fact, this occurs, although less directly, in writing as well- the writer tailors his writing style for the presumed audience, for example, as seen in the contrasting styles between a professional text and a casual note to a friend.

Since language communication, be it written or oral, is not conducted in a vacuum, if the goal of a speaker is to convey information effectively to a listener, one cannot overlook the relationship between speaker and listener in the study of language production. When communication is one-on-one, the speaker/writer may more carefully tailor his words to his audience, according to how well they know each other, and what their relationship is; when communication is to a group of familiar or unfamiliar people, expectations and presumptions of common features among them are, or should be,

decreased and adjusted accordingly.

In communication, the participants must make use of their common ground- that is, what is known to be understood by both of them. If the involved parties know little about each other, they presume their common ground to be relatively little, and will provide a greater amount of information. Even between strangers there can be presumed to be certain things in common simply because they are both humans, adults, Americans, New Yorkers, perhaps even from the same neighborhood. Similarly, two people of different nationalities may find a lot in common because they are in the same line of work, and can subsequently speak like experts, at least on their common topic (e.g., Isaacs & Clark, 1987).

Several studies have focussed upon and emphasized the need for creation of common ground and collaboration of the dyad in conversation, or in creation of any reference or description (e.g., Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987; Krauss & Glucksberg, 1977; Schober & Clark, 1989). As common ground is established, the number of words needed to describe the stimulus decreases. Clark and Wilkes-Gibbs (1986) had a speaker describe unconventional figures consisting of combinations of geometric shapes that were arranged in a particular sequence to a listener who had the same figures but in a random sequence, and studied how the pairs collaborated, in successive trials, to arrive at the correct arrangement of figures. Efficiency of communication was measured by the number of words used in each repetition of the task, and by how much turn-taking between the pair was required to successfully arrange the figures. Be it through defining of the referent over the trials (e.g., Krauss & Weinheimer,

1967; Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989) or because the involved parties learn what their partner needs for successful communication, common ground is established, and the number of words used lessens with each trial. In a similar task, Krauss and Glucksberg (1977) found that while adults could perform this task correctly and easily, children could not effectively communicate the material; they used cryptic or idiosyncratic descriptions of the figures that the hearer could not comprehend. The authors speculated that the children could not yet use language to develop a socially shared code. These results suggest that this ability is a learned one, or one that requires a level of cognitive development and understanding to relate to the perspective and needs of others.

Using the stimuli of Clark & Wilkes-Gibbs (1986), Schober and Clark (1989) studied how, in the building of common ground, active collaboration between the participants facilitates more successful communication. They compared the performance of the active participants in a matching task to that of someone overhearing the dialogue. As predicted, they found that even when the overhearer shared the same background knowledge as the participants, he was less able to perform the task as correctly as the active collaborators. Furthermore, overhearers who heard only later communications between the participants performed even more poorly than those who overheard the entire dialogue. The findings differentiating between participants and overhearers provide additional support for this notion of specific tailoring of information for an audience. (However, these studies did not focus upon the possibility that dyads of same and mixed sexes might interact differently from each other, although Schober and Clark [1989] did

present evidence that some dyads communicated more effectively than others.)

Similarly, Isaacs & Clark (1987) compared the establishing of reference by experts and novices, and the differences in their need to assess, supply, and acquire expertise. As the parties assessed each other's knowledge, they modified their style accordingly, the speaker providing more information if necessary, or using jargon unfamiliar to any but a sophisticated audience, the listener asking for further clarification, or conversely, demonstrating familiarity with the material and thus requiring less information.

That adults take into account whom their intended audience will be in creating their descriptions even in indirect communications such as writing was demonstrated by Fussell & Krauss (1989), who showed subjects abstract, odd shapes, and had them write descriptions for later identification by 1) themselves or 2) someone else. The assumption was that in descriptions for oneself, common ground was total, and thus would differ from communication with another person, with whom common ground was less than absolute. They found that the descriptions differed in terms of number of words, as well as in usage of figurative language. Descriptions for another person consisted of a larger number of literal references, such as geometric shapes, which are generally believed to be common to everyone, while for one's self, descriptions consisted of fewer words, and were more idiosyncratic. An advantage of this study was that while it did not require direct interaction between subjects, the communicative goal was made explicit to the subject, yet without the possibly confounding dynamics of a dyad.

While the studies described above emphasized the notion of commonality

between the communicating parties, and the development of reference through dyadic discourse, Hupet, et al. (1991) focused upon a different aspect of the communication: the complexity of the referent itself. Using an expanded set of the stimuli used by Clark & Wilkes-Gibbs (1986), their results mirrored those of those authors' in that as the definite referent was established, less feedback was required, but also demonstrated that greater collaboration was needed when referring to items that were very similar, versus those that differed in obvious ways. Similarly, Krauss & Weinheimer (1967) found that dissimilar stimuli were more easily distinguishable (more "codable") than were stimuli that were very similar.

Common ground can thus be described as a continuum of familiarity between the communicating parties: At one end of the spectrum are people who fail to ascertain each other's communication needs, or in the case of writing, fail to take the needs of the presumed audience into account, and the result is inefficient or even ineffective transmission of information. Failure to establish common ground severely limits the potential of communication even when it does not end communication outright. At the other end are people who maximize their common ground, often through cooperation that allows the development of ideas in an infinite variety of ways, or through utilization of language in ways that enhance comprehension. For example, by turning to creative, nonliteral language, a speaker can successfully go beyond the standard range of a word's meaning and utilize novel forms of language that, while perhaps unintelligible or nonsensical out of context, in the situation at hand express fully and clearly the concept under discussion (e.g., Gerrig & Gibbs, 1988).

Thus we see that decisions are continually being made by the speaker, who takes into account what he knows about the listener before choosing what he believes to be the optimal way of expressing his thoughts. Internal, unconscious processes also determine how the speaker arrives at his conclusions. They impact not only upon how a speaker perceives a situation, but how he chooses to express this perception. The selection of words and phrases and their usage in the construction and production of language is not an arbitrary or simple process- what one speaker finds communicable through one style, another person might find through another: expressibility then, is more than an objective situation-- it is a situation as seen by an individual. The individual, for whatever reasons, has chosen to resort either to literal or to nonliteral language in order to *best* convey his intentions.

His choice may be conscious and deliberate, or it may be unconscious, due to psychological or personality factors. It is also intrinsically limited by his cognitive abilities and knowledge. For example, it is possible that someone with very sophisticated language abilities may be more likely to utilize language in creative and innovative ways, while someone with an impoverished lexicon would be limited in their ability to express themselves as efficiently, and, lacking a particular word, might therefore be more circumlocutory. Alternatively, someone with a weak vocabulary might find himself forced to use the words at his command more creatively than someone with a more replete lexicon. However, the study of cognitive variables and production of figurative, or creative, language remains relatively neglected. One early study on cognition and figurative language (Pickens & Pollio, 1979) looked at the correlation between adults'

patterns of performance on several tasks believed to be related to aspects of figurative language co- petence. The authors concluded that there is a great deal of variability between subjects, as well as within-person consistency, in terms of preference and comprehension of nonliteral figures of speech. They stressed, though, that beyond individual competence is the effect of context and the specific task , which serve to influence and play a role in a person's ability and preference in communication. However, the authors did suggest that it may be possible to identify individuals who are differentially sensitive to figurative expression. In a study of simile comprehension in children, Readence et al. (1983) found that, not surprisingly, vocabulary or lack of it, might contribute to a failure to comprehend metaphorical language, in that a knowledge of the salient matching features of the simile must be present. O'Brien et al. (1986) found this same impact of vocabulary in adults, as well. It stands to reason that this intellectual variable will also impact a speaker's *production* of language. To that end, in the study below, we will include an analysis of the cognitive/intellectual function of vocabulary, in an attempt to determine whether and how it contributes to literal and figurative language production.

LEXICAL ACCESS AND SELECTION

An ability necessarily correlated with vocabulary that also impacts upon language production is that of naming, the invoking of a particular word. A person must not only have a word in his lexicon, but he must be able to access it on demand, i.e., name it. The failure to access a correct or desired word can be attributed to different causes: to a lack of knowledge of the word in question, i.e., a deficit at the conceptual level, or to an

anomia, an inability to produce the name. In the case of certain aphasias, the patient may be unable to access certain words, or produce an unintended word, or mispronounce the intended word, and may not be aware of his error. This probably differs from the instance of someone whose syntactic system is intact, yet who fails to produce the word upon demand, be it his own desire or someone else's. Often the subject is able to elaborate on certain properties of the word, yet remains blocked as to the word itself. This latter phenomenon has been called the tip-of-the-tongue phenomenon (TOT) (Brown & McNeill, 1966; Burke, et al., 1991). It has been proposed (Burke et al., 1991) that a TOT occurs when connections to some or all of the nodes weaken to the point where they fail to prime sufficiently to enable activation. A network model of activation can be used to describe how the word is accessed (Bock & Levelt, 1994). The working assumption is that the picture activates the concept in the brain; the concept node is activated, and the activation spreads down from the conceptual level to the lemma level (the word as meaning and syntactic entity) then to the lexeme (sound) level. In more complicated language production, as in complex sentences or grammatical structures, not all words correspond to concepts (e.g., prepositions, or words that must be used in conjunction with other words) and they must be activated indirectly.

However, evidence suggests that the abilities of vocabulary and naming, while necessarily highly linked functionally, are representative of separable cognitive functions. One realm in which the separation between naming/production and vocabulary/comprehension is more clearly evident than usual is that seen in normal aging. Many older people both complain of, as well as demonstrate, increased difficulty in

naming; however, there appears to be a differential rate of decline between that and a relatively preserved verbal ability. Burke & Laver (1990) propose that the preserved ability in older people is in comprehension, and the complaints are in production. Others cite differences between new learning vs. old; fluid vs. crystallized; episodic vs. semantic. According to Node Structure Theory (Burke & McKay, 1991) production involves the top-down spread of semantic priming to phonological nodes, while comprehension involves the activation of phonological nodes and the bottom -up spread to the semantic nodes. In their model, activation is necessary to retrieve the information each node contains, is all or none, and does not spread. Priming prepares a node for activation, and an activated node primes all those nodes connected to it. The strength of the linkage, i.e., the strength of the connections determines the rate and amount of the priming between them. Declines are seen in old age due to a decline in linkage strength, or possibly due to a potentially longer interval of disuse. According to this model, priming *diverges* from a lexical node to many phonological nodes, and *converges* from semantic nodes onto a single lexical node. A deficit in priming would have a greater effect when it diverges than when it converges, thus explaining the differential decline in comprehension and production, and how the two different processes may each be affected differently by age (Burke & Laver, 1990). While a word must exist in one's lexicon in order to be used (vocabulary), the word must also be available on demand (naming). The frequency of a word's usage contributes to its accessibility, since its links to other words and concepts are strengthened with use, and its nodes are more easily activated.

That naming and vocabulary represent different cognitive functions leads one to

the question of how a deficit, even a relative deficit, in one or the other would affect language production. As yet, this question of whether and how differing intellectual functions such as naming and vocabulary scores would manifest in production style (e.g., phrase count and figurative/literal intention), has not been systematically studied. Furthermore, current methodology may be ill-equipped to distinguish subtle differences of a normal population: while priming, reaction time, and semantic inhibition paradigms are often used in naming tasks, spontaneous picture naming is generally not studied except in the cases of disability, so most measures of naming are not geared towards a normal population. Thus the risk exists of a ceiling effect, and subsequent loss of sensitivity, when testing normals with a test geared predominantly towards those with impaired language function. Despite these limitations, included herein will be a preliminary effort to study the contribution of these intellectual strengths and weaknesses to language production.

LITERAL AND NONLITERAL LANGUAGE

The relationship of the speaker/listener interaction in communication has been discussed above in terms of creation of reference and word selection, and in the interrelationship of certain intellectual functions. These factors interact to affect how language is produced and understood. One way of studying language is in terms of the words that were used to represent the intended meaning. Every word has its standard, accepted, "dictionary" meanings, about which there is general agreement. These are often considered its literal meanings. In addition, there is figurative meaning, in which words are used to communicate new insights or alternate meanings, often by means of

comparison with something in a different category (Bartel, 1983), or through their use in a situation where their literal meaning is false or inappropriate. Examples of figurative language include metaphor, and simile, and others, which assume figurative properties only in selected contexts, such as indirect requests and metonymy.

Early research did not directly address meaning in terms of communicative intention, but rather focused upon structural properties of isolated words as part of a semantic system and semantic access in memory (e.g., Collins & Quillian, 1969). For example, Katz & Fodor (1963) proposed that the lexicon consists of words that share certain features of meanings, for example, "male" is shared by "bull" "policeman" "quarterback", etc., but also contain distinguishing properties that make each word unique (E.g., for "uncle", the particular property is the kinship relationship.) These semantic systems and the memory studies derived from them (e.g., Collins & Quillian, 1969) by and large do not account for the effect of contextual and extralinguistic features upon language production (Tartter, 1986). However, the idea of selecting specific features does lead into the realm of context-driven meaning, in which properties in language are used to create new or different meanings.

What makes language literal or figurative has been the subject of debate, both philosophical and linguistic, throughout history. On one end of the continuum are proponents of the belief that all thought may be more properly reflected in figurative language than in literal. For example, Lakoff & Johnson (1980) have suggested that our conceptual system is fundamentally metaphorical in nature and thus, even *literal* language is metaphorical, in that the human conceptual system and thought processes are

metaphorical. In other words, concepts are often understood in terms of other concepts such as terms of direction or orientation, as in "I was feeling up" (GOOD IS UP) (p16) or in terms of entities and substances, as in "He ran out of ideas" (IDEAS ARE RESOURCES) (p 48). They have tried to demonstrate that concepts are at least partly structured by the metaphors in everyday language, which are based on human interaction with physical phenomena. (However, many of these concepts are considered by others to be literal, in that they are static and universal.) In this light, all language can be seen as reflecting and building upon preestablished concepts, and thus can be considered fundamentally figurative in nature.

At the other end of the spectrum was the older, traditional view of language, that literal language is that which is understood independently of context, through the meanings of its component words, while figurative language belies its literal meaning in the context (as reviewed and critiqued by Gibbs, 1984). Early researchers supported the hypothesis that in terms of language comprehension, the literal meanings must always precede any alternative (nonliteral) meanings which may actually be the intention of the speaker; only after the literal meaning has been tested and eliminated for the particular context, is the figurative alternative tested.

This view of figurative language comprehension as necessarily secondary to the literal is not supported by research (e.g., Clark & Gerrig, 1983; Gibbs, 1984; Gildea & Glucksberg, 1983; Glucksberg, 1989); in fact, it has been demonstrated that in certain situations, specific types of nonliteral language, such as idioms, metaphors and indirect requests, are understood automatically and the alternative but inappropriate literal

meaning is bypassed. For example, in the indirect request, "Where's the salt?", the conventional and (usually) correct response is to pass the salt, not answer the question. That context plays a role in directing which meaning, the literal or the figurative, is intended, has been seen experimentally as well (Glucksberg, 1989; Glucksberg, Kreuz and Rho, 1986). In a lexical decision paradigm, subjects presented with polysemous (words with two or more relatively common meanings) primes only responded more quickly to those primes that were contextually appropriate, regardless of whether it was technically literal or figurative, while the potentially ambiguous words that were not appropriate to the situation were responded to like completely unrelated targets. That certain types of figurative language forms are understood necessarily, at times before the literal, speaks against the probability that they are a cognitively abnormal process.

Some (e.g., MacCormac, 1985) have suggested that Lakoff & Johnson have simply changed the linguistic precedence from literal/metaphorical to that of metaphorical/literal, and that by their insistence that most language is metaphorical, they have become forced to describe literal language in metaphorical terms, which is even more difficult than the opposite process of describing figurative language in literal terms.

To MacCormac (1985), who has adopted a view in-between the proponents of either literal or figurative language predominating, literal can be defined as the use of ordinary language to express concrete things. The line between literal and metaphorical is not clearly drawn, since metaphors are in continual transition to becoming ordinary language, and only familiarity and context determine how ordinary language is; for example, the innovation that led to the idiom "raining cats and dogs" has lost its

creativity, resulting in its current frozen form. Burbules, Schraw & Trathen (1989), while endorsing Lakoff & Johnson's view that metaphorical language is not deviant, claimed that MacCormac's ideas were more similar to their own, by their contention that literal meaning is a reflection of the sentence within a specific context. Burbules, et al., extended this into the thesis that literal and metaphorical interpretations are degrees of the same process. Olson (1988) further suggested that only if a statement has two or more possible interpretations, and one can be recognized as metaphorical, then the one closest to the wording of the sentence is taken as the literal one. Thus, both literal and metaphorical meanings require an awareness that alternate meanings from that actually intended by the speaker are possible. Without that understanding, the sentence is neither literal nor figurative: it is simply understood. That is not to say that the literal meaning must first be comprehended and then rejected in favor of the nonliteral; rather, each may have degrees of automatic understanding, according to the context.

Based on evidence that the context can affect or determine the meaning of a word or phrase (e.g., Gibbs, 1984), Gibbs, too, minimized the difference between literal and metaphorical language, claiming that there is no "principled distinction". Literal and figurative meanings may be different interpretations, but he claimed, not necessarily because of different interpretive processes. In fact, the processing method is the same for both literal and figurative meanings, that differences are "more a matter of degree than kind... To speak of a sentence's literal meaning is to already have read it in light of some purpose, to have engaged in an interpretation. What often appears to be the literal meaning of a sentence is just an occasion-specific meaning where the context is so widely

shared that there doesn't seem to be a context at all" (p296). Thus, while perhaps in a "null" context, literal meaning might always predominate, in life there is always some sort of context, based on the situation at hand, the relationship of speaker and listener, and the like. The model of Gibbs suggests that words do not have specific core meanings (e.g., MacWhinney, 1989), but instead compete within a semantic niche in a continuous decision-making process: On different occasions, a word will be retrieved for different of its attributes. For example, at times "elephant" will be retrieved for its representation of a mammal of Indian or African origin, while at other moments, for its property of immense size, and at still others, for its reputed characteristic of long memory. Concepts, then, are temporary representations rooted in a particular context, and selected from long term memory, which itself contains deeper and additional properties to the ones selected. Thus, in one context, a word may utilize certain of its properties, while in another, other of its characteristics will predominate. And according to the properties emphasized, these words might be considered literal in one situation, and considered nonliteral/figurative in another. That the distinction between figurative and literal itself appears to be context-dependent complicates and perhaps obviates any definitive study of figurative language, and this futility perhaps accounts for much of the seeming directionlessness of the research.

Yet this very lack of consensus permits the freedom to define the terms of study, stimulates thought and discussion, and in fact allows us to elaborate and expand upon the terms under study. In examining the polarity of the literal/figurative debate, it appears that in a certain sense the poles meet, and differ most vastly only in the terminology used.

Most researchers seem to agree at some level that a gradient exists of figurative and literal meanings, and that context and intention must be considered in order to determine where on the continuum a meaning lies. In a more global sense then, one that may perhaps be more suitable for the purposes of content analysis, the term "figurative language" can be extended beyond production of the various figures, such as metaphor and analogy, to include language whose *intention* is to describe what a figure "is like", regardless of whether it is a standard form such as metaphor, or through creative placement and/or usage of words in a new context. Literal language on the other hand, describes, usually analytically, what a figure "is", using directions and geometrical terms as references, since these terms require little or no understanding of common ground, and are considered universal (e.g., Fussell & Krauss, 1989).

While comprehension of figurative material appears to be nonoptional and automatic under certain contextual conditions, (e.g., Gildea & Glucksberg, 1983), the conditions causing the *production* of figurative language are less understood than those of comprehension, so that it remains unclear what the processes are that caused the speaker to utilize these speech forms in the first place. Whether they are simply more literary, flowery ways of saying what could be said literally, or whether they serve an additional purpose, necessary for successful or more effective communication in certain situations, has been the source of recent discussion. It has been suggested that analogies are used to describe something unfamiliar or unknown to one's audience, or if a concept is unfamiliar or abstract. An analogy, in the form of a metaphor, a simile, etc., would liken the unfamiliar to something within the experience of the listener (Glucksberg, 1989).

As has been iterated above, the primary goal of a speaker is to convey most accurately his/her intended meanings. If creative language is used, presumably there was a reason, and it may be because it is necessary for the successful expression of something that is otherwise uncommunicable, or less effectively communicable. Thus the use of any form of creative language could imply that some concept or idea was *inexpressible* within the confines of standard meaning. In this case, the notion of inexpressibility need not mean a literal term is completely nonexistent or unavailable, but rather that it is not selected by the speaker, presumably because it has been deemed inadequate (Gerrig & Gibbs, 1988).

The issue of what constitutes communicability/expressibility has been studied recently by several researchers. For example, Fainsilber & Ortony (1987) discussed three communicative functions of metaphor/figurative language: to communicate that which is difficult or impossible to express via literal language (*inexpressibility*); to convey a great deal of information more succinctly (*compactness*); to capture the nuances of a subjective experience in order to better convey it (*vividness*). Theirs was a first step toward establishing that some certain things/situations whose descriptions regularly invoke more figurative language (specifically metaphor) than others- i.e., figurative language is more necessary in some situations than in others. Subjects were asked to describe, in writing, highly and mildly emotional episodes in their lives, and to describe their accompanying actions/responses to the situations.

In their study, Fainsilber & Ortony focused upon the features of vividness and expressibility, using emotion as a vehicle to elicit metaphor. They hypothesized that

people would describe emotional, subjective experiences more figuratively than they would describe activities or behaviors, and that powerful emotion would elicit more metaphor than would weaker emotion. In fact, although their results yielded an overall low output of figurative language, the results suggested that more intense emotion did lead to increased usage relative to situations of weaker emotional import, and that emotional situations promoted increased production relative to the description of actions.

As has already been mentioned, ease of describability/codability of the stimulus in question leads to different amounts of language output (e.g., Hupet, et al., 1991; Krauss & Weinheimer, 1967). Content of the output has only been studied slightly (Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1989) and even then, focus has been upon the relationship between speaker and listener, not in terms of degree of expressibility of the stimulus, i.e., the relationship of the speaker to the stimulus. Whether the content of the output differs in this situation has yet to be ascertained. In fact, content has not been studied extensively, except in terms of word count, verb/noun ratios, and type-token ratio. While these measures may be valuable in terms of providing information about word frequencies, they are highly susceptible to individual stylistic differences, and do not provide information about content, in terms of imageability and meaning of the word choices.

The third purpose of metaphor proposed by Fussell & Krauss was that of compactness. This aspect, brevity, has been studied as a secondary issue by researchers studying reference (e.g., Clark & Wilkes-Gibbs, 1986; Krauss & Glucksberg, 1977), and addressed primarily by Ford (1980). In this study, same-sex pairs of subjects were asked

to solve a problem, some using as few words as possible, while other pairs were not given that latter request. The author looked at the parts of speech generated, but not how the words were intended, i.e., not the literal vs. figurative content of the output. And although content was not the primary focus of the "reference" studies, it was found that as the referent was established between communicating parties-- and quite often this referent was figurative-- the number of words needed to describe the stimulus decreased. The referent often consisted of a holistic description of abstract shapes. As the salient features were selected by the speaker, segmental geometric reference decreased, and holistic, global reference increased (Clark & Wilkes-Gibbs, 1986).

WRITTEN VS. ORAL MODALITIES OF PRODUCTION

It is obvious that speech and writing differ. While each performs the act of linguistic communication, they do so differently, both in terms of output as well as in their intrinsic properties. Some of the stylistic differences between written and spoken language have been described and discussed in a review by Chafe & Tannen (1987). They cited studies which found that written language used shorter text, longer and more difficult words, and a more varied vocabulary, while spoken language contained more repetition and elaboration and more self-reference. In a more general sense, the modalities differ in some fundamental ways, most obviously in terms of immediacy of the production. As mentioned earlier, speech and thought occur virtually simultaneously, while written language has the luxury of delay and planning. Speech is acquired without formal instruction, while writing has to be consciously learned; it relies on systemized procedures, and children are taught to be prescriptive, to select words and phrases

carefully, to arrange them, and to abide by a multitude of rules. Thus it is more deliberate and contrived than speech. In addition, it is artificial, in that it requires manipulation of a tool, while speech is natural, using only speech organs (Akinnaso, 1982).

Furthermore, in oral production, while the speaker must try to tailor his production to the listener's immediate needs, he can respond and modify his output according to feedback by the listener, can repeat as necessary, and can use extralinguistic and nonlinguistic cues like prosody and facial expression for emphasis or to develop nonliteral meanings such as irony. In contrast, written language is generally presented indirectly, i.e., there is no immediate response from the listener. This allows the producer to plan his output, but this lack of immediacy also prevents his utilizing any additional, nonverbal cues. The response of the recipient is less immediate if present at all, and so the writer is forced to tailor his output to the *perceived* needs of the presumed recipient. This lack of feedback in such noninteractive situations is less fluid than a dialogue, but it enables us to focus upon the perceptions of the producer, who must make his own assumptions concerning common ground, rather than upon the specific dynamic between the communicators. Tannen (1982) makes the distinction that oral production makes maximal use of context, while written production must be explicit, and rely on lexicalization, i.e., the words themselves, in order to establish cohesion.

These differences may account for, or contribute to, some of the stylistic differences between written and spoken language. Although many studies cite multiple differences, few of these studies consist of research done using parallel tasks between the modalities, and it has been suggested that some of the wildly varying findings are due to

inconsistency in the methods for studying the production (e.g., Akinnaso, 1982; Chafe & Tannen, 1987). Furthermore, while nouns, verbs, and other parts of speech were discussed, the intentions of language, i.e., the content, whether literal or figurative, was not elaborated upon. This is a notable omission given the importance of the metaphor/literal distinction in understanding language usage. A methodical study of how figurative language is used in writing and speech, and whether and how it differs between the modalities may lead us to an understanding of the importance of assumed feedback and intended audience in word choice and preferred modes of communication. Having accounted for the relative brevity and lack of repetition in written language in comparison to spoken, we are as yet unclear whether the style differs- for example, whether people will be more literal in the inherently more formal writing than in speech, in a situation where the task demands between the two are not a confounding factor. In the study described below, we will do a comparison between speech and writing styles, which we hope will more successfully reflect the inherent differences between these modalities than earlier studies which were not specifically designed to do so and studied differences only post-hoc.

AGE AND SEX FACTORS

Other factors such as age and sex may contribute to different language styles. For example, over the lifespan, normal changes in both written and spoken language have been documented (e.g., Bromley, 1991; Pratt, 1991; Walker, et al., 1988). The effects of normal aging upon naming have already been discussed above; in more complex tasks, such as discourse, changes have been seen as well. For example, in a writing task,

Bromley (1988) found a slight decline with age in sentence complexity, but that overall, language vocabulary functions remain stable with age and may even slightly improve. Interestingly, while men and women did not demonstrate significant differences in most areas, women's writing was considered "more readable" (shorter sentences and/or words) than the men's. The authors note that the variables of word length, use of long words, readability and word output were more a reflection of vocabulary and educational status than of age or sex. Similarly, Kemper, et al. (1989) found that more education affected oral and written language in terms of number of words used, but memory capacity affected sentence complexity even among less-well-educated persons.

In a study of oral discourse in women of varying ages, Walker, et al. (1988) found age differences as well; like Bromley, the older women used more variable vocabulary, but younger subjects used longer utterances of greater syntactic complexity. Of interest was a stylistic difference, in that older women appeared to use more repetitions, interjections and revisions as they spoke. The authors speculated that this production style may appear to a listener to obscure the accuracy of the output, even though it is in fact, correct. In contrast, Kemper, et al. (1989) found that while older subjects produced less complex sentences, possibly due to reduced memory capacity, their language samples were judged to be clearer and more interesting than those of the younger subjects. Similarly, Pratt & Robins (1991) found that the oral narrative styles of older adults were judged to be of better quality than that of a younger group, and further, that the discourse samples of the subjects with the highest vocabulary scores (WAIS-R) were preferred over those with lower scores. The authors note, however, the limitations of a cross-sectional

study, in terms of the possible age and cohort effects; for example, oral narrative may have been emphasized and encouraged differentially at different points in history, and this would be reflected in the superior performance of one group over another, for whom such a skill was less valued. Kemper, et al. (1989) suggest that there is a trade-off among the language skills with increased age. While some language skills decline with age, others improve.

Differences in language production have also been seen between males and females. For example, Tannen (1990) has described extensively the different conversational styles used by men and women in everyday situations. Bilous & Krauss (1988) looked specifically at nonlexical differences in style when men and women interact in same-gender and mixed gender dyads. All subjects participated in both dyad types, and were asked to resolve a problem between themselves. It was found that males' speech productivity did not vary as a function of their speech partner, while women's did, decreasing (to the same amount as the men) in a mixed dyad. Overall, they found differences in some variables, but not in others, and emphasize the complexity and difficulty of generalizing any effects of dyad interaction. Mulac & Lundell (1986) attempted to identify clusters of features that contribute to the overall effect of gender. They performed a study involving subjects ranging from adolescence through adulthood describing photographs orally, and found a number of gender-distinguishing variables. One study de-emphasized the differences: Pillon, et al. (1992) found that in an experimental setting involving (French) male and female dyads, strangers, men's and women's conversational behaviors were similar in terms of degree of participation, such

as number of words per utterance and turn-taking. Although differences between the men and women were in fact present in this sample, they were minimized in favor of the similarities. This latter study's findings, or rather, its interpretation of the findings, seem to be the exception; more commonly, irrespective of the testing conditions, differences in style between men and women have been documented. It must be remembered that the varying methodologies may obscure or emphasize differences or similarities that may or may not be artifacts, but that appear in the conclusions drawn from the research.

While overall, there appears to be acknowledgment of the differences in linguistic style between men and women, measured in terms of variables such as pauses, fillers, and sentence length (e.g., Mulac, 1986), as in other areas of language production, there is a paucity of research on the content of the language sample- once again, there is a surprising lack of research on style and intention of the production- namely, use of metaphor, and more generally, nonliteral language.

It is possible that this gap in research is due to a lack of metaphor in natural language, but based on the supposed prevalence of metaphor (e.g., Pollio, et al., 1990) and of metaphorically used language (e.g., Lakoff & Johnson, 1980), it is more likely that this area has simply been overlooked. For example, the above studies that looked at the creation of reference focus primarily upon the establishing of common ground between communicators, but generally ignore the creation of novel combinations, and utilization of standard words in novel contexts, to do so. The above "metaphor" studies focus predominantly upon the comprehension of different types of language, figurative and literal, or more accurately, denotative and connotative to their varying degrees. However,

they do not focus upon the conditions of their production. Furthermore, because there is a lack of consensus in the terminology being used, different studies that purport to be exploring figurative language are often studying different phenomena.

This dissertation is a study of figurative language, necessary because figurative language is an important element of language whose purposes have been long misunderstood. In addition, this is a study about the *production* of figurative language, since production of language in general, and of nonliteral language in particular, has been relatively neglected in comparison to language comprehension. More generally, it is an exploration of word selection and context, and how different linguistic choices might reflect internal and external motivating factors. Finally, this study attempts to place language into a conversational/constructionist interactive framework that develops the conditions that make people elect to use one word over another, and nonliteral language over literal, and furthermore, that allows the speaker's intentions to play a major role in determining whether the word used is figurative or literal.

METHOD

PARADIGM AND STIMULUS SELECTION

This dissertation is aimed at studying the production of figurative language. To do this, we needed to use a task in which subjects were likely to generate figurative language. Some preliminary studies found this to be not so easily achieved. This section describes what methods were attempted and what methods were ultimately selected.

In an initial attempt to study figurative language in natural production, for several weeks I listened informally to dialogue in everyday life, and was attentive to any

metaphors or figurative usages that I overheard. Surprisingly, considering the prevalence of nonliteral figures of speech documented in daily language by various researchers at rates ranging from 4.9% to 0.6% depending upon the criteria used for assessing the sample, and the language sample used (e.g., Pollio, et al., 1990), I heard very few instances of spontaneous production of metaphor in normal conversation among individuals. Notwithstanding that in the Lakoff-ian sense of metaphor, virtually *all* words are metaphorical, which is the view that Pollio, et al. (1990) factored into the amount of usage, nonetheless, little language was actually being *used* metaphorically, i.e., in ways designed to mean different things than what their dictionary meanings provided for. Hearing so little of what I was listening for led me to realize I had to *elicit* the production of figurative language in order to get language samples in any meaningful amount, as well as to extend the realm of figurative language from metaphor and idiom to include analogical and creative language.

In a first attempt to have subjects generate figurative language, I attempted to replicate the study of Fainsilber & Ortony (1987), who had subjects describe emotional situations. Although their yield had been low, theirs was one of the few studies that attempted to elicit figurative language in any systematic way. As described above, these authors had postulated that inexpressibility, vividness and compactness were among the communicative functions of nonliteral language, and they had designed a task using emotion as a means of eliciting language.

Because their overall yield of metaphor pointed to increased intensity of emotion as a source of increased metaphor production but was low in terms of total output, I

decided to informally replicate their study with modifications. The task was performed orally instead of in writing, as the authors had originally done, using the rationale that perhaps the change of modality to the more spontaneous oral mode might result in greater metaphor use, on a reduced scale (n=8) and with a smaller set of emotions (2 per person, one positive, one negative) so as not to tire or overwhelm the subjects. Like Fainsilber & Ortony, these results showed overall poor yield of metaphors, (at a rate of only 1.4 metaphors per description, range 0-3) and consistent with the results of those authors, increased intensity of emotion generated more figurative usages (mild emotion generated none), and emotion generated more figures than did description of their activities. It was found that the personal style of the respondent appeared to contribute substantially to their style of output. The subject variability assumed several forms: that of their selection of an incident to recount and that of their idiosyncratic language style. It was suggested that a more structured task would eliminate some of the subject variability found in open-ended discourse. In addition, providing specific stimuli designed to be ambiguous might elicit from subjects figurative language in higher yield according to the describability of the stimulus in question.

To that end, in an informal pilot study (n=6, ages 10-53), I replicated the methods of Fussell & Krauss' (1989) written description task, in which subjects were asked to identify abstract figures for later re-identification either by themselves or by someone else. In place of Fussell & Krauss' original set of stimuli, due at first to difficulties acquiring the full set of the stimuli, and then to difficulties in linguistic coding that had been expressed by the authors, I used the assorted figures from the Chinese game of

Tangrams used initially by Clark & Wilkes-Gibbs (1986). These figures consist of seven geometric shapes that are combined in various forms to create figures which resemble, to greater and lesser degrees, people, animals and birds. These stimuli were chosen deliberately for their resemblance to natural objects, and for their geometric properties; i.e., the figures consisted of features that could be described either analogically or literally, according to which features the describer felt were more salient in a particular situation.

As mentioned earlier, these stimuli have been used in a number of experimental paradigms studying the development of reference and dyadic communication. The focus of those studies was primarily on aspects of interactive communication and the development of reference over several trials, and the descriptions generated showed a range of figurative and literal responses, based largely on how solidly reference was established. Generally, in earlier matching trials, more literal terms were employed, while in later trials, more holistic or figurative terms were accepted by both parties. Specifically because of this usage of both types of language, in the current study, in which the focus was primarily upon the content of the descriptions, we felt this set of stimuli was well-suited. It would generate a consistent amount of figurative and literal responses which were seen to vary with familiarity. It was hoped that the results would mirror those of the initial matching trial of other researchers. Instead of manipulating the common ground, however, we would vary the audience condition in order to study the degree to which descriptions would change with different audiences. In addition to studying the amounts of figurative and literal language, we also introduced a measure of

hierarchical categorization, in which the level (superordinate, subordinate, and distinguisher [Katz & Fodor, 1963; Collins & Quillian, 1969]) and specificity of each description would be examined in each condition. It is believed that the selection of one style of description over another would reflect the producer's own perceptions of the stimuli vs. her impression of another person's perceptions, thus measuring a person's confidence in their own mapping process.

In the current task, an expanded set of stimuli was used, and subjects were asked to describe each figure for later identification in *both* ways, i.e., 1) by themselves and 2) by others, rather than in only one condition (e.g., Fussell & Krauss, 1989). We believed the comparison of the same subject's responses between conditions to be a more accurate way to measure how an individual's style changes with their perception of a situation, and therefore to reflect their personal understanding of the task . Subjects responded easily to this modification, and it was therefore incorporated into the protocol, so that all subjects received all test conditions, both written and oral, for both audience conditions. However, it was found necessary to separate the two task demands, such that subjects responded first to the full set for one audience condition, and then for the second condition: when asked to do one then the other consecutively for each individual stimulus, subjects appeared to keep the same description, either in its entirety, denying there would be any differences between conditions, or just adding one or two words elaboration. In addition, it was found necessary at the start of the task to present the subjects with the full set of stimuli, in order to orient them as to the demands of the task, in terms of the similarity between the stimuli. However, all stimuli could not be visible at all times- when they

were, subjects tended to assign labels like 'the third one' or 'the little one' instead of creating referential descriptions.

Both men and women of varying ages were included in a limited sample (n=8, ages 16-26 years) which was conducted both orally and in writing. At the end of the different audience tasks, subjects were asked to describe some of the figures literally, if they had not already done so, in order to obtain a sample of their literal language to compare its style and degree of describability to their spontaneously generated figurative samples. Although not statistically analyzed, (the results would not have approached significance due to the small sample size,) it appeared at least superficially that the men and women demonstrated stylistic differences, in that the men used fewer full sentences, in favor of bullet-style descriptions, while the women tended to prefer a fluid discourse style. In addition, the population varied in age, and therefore in the era in which they were educated, and subsequent production styles. It was decided that in order to minimize any possible variability based on these possible contributing factors, to limit the test population to a single sex (female), and a particular age range, i.e., to a single cohort.

PURPOSE & DESIGN

The purpose of the following studies was to investigate the effect of different contexts (audience conditions- self and other) upon the type of language output (literal/ nonliteral) in the description of abstract figures generated in both written and oral modalities. The effect of competence in naming and vocabulary was explored in conjunction with amount and style (preference for literal or nonliteral) of the language output. The effectiveness of the descriptions (number correctly identified) was then

judged between the different audience conditions. Only women were included in this sample population, in order to eliminate variability of sex. Likewise, only one age group was selected, as it has been shown that linguistic style may vary with age. Context was limited to the description of complex stimuli, with features of common ground pre-established by the task demand, rather than through interaction between speaker and listener, since results have shown differences in production in dyads with differing interactive skills (Schober & Clark, 1989). By using specific stimuli rather than an analysis of spontaneous dialogue or a description, we are creating a specific target for the producer. In this way we can limit the variability of selection of material. Likewise, the context was specified for each modality of production, so that any differences seen between audience conditions would be a reflection of the differences of the modality. By controlling the task in such a manner, differences in production could be studied in identical contexts, enabling us to learn, in a systematic way, more about what factors contribute to styles and stylistic preferences. While defining the context necessarily limits the scope, we believe that it nonetheless approaches the study of figurative language in the real world. Although strict metaphor is not being elicited, we contend that in the sense that language whose intention is analogical is being compared with that which is literal, the language being produced is in fact figurative. A novel construction is not being created, except insofar as a representation is being created from shapes into a whole under particular conditions that appear to elicit this effect. In fact, we assert that that *is* figurative language in a natural setting: words or phrases that take on new meanings or usages according to context.

It is hoped that an examination of the contextual conditions under which these subjects elect to use nonliteral language will lead to increased knowledge about the purposes and functions of figurative language in description and about the producer's perceptions along with external conditions that lead to their usage. While it is unclear to what degree these results can be generalized to other populations that vary in education, age, or in cognitive ability, this study at least can be considered a first step in creating a methodology for examining different styles of output, and how words are used to extend or constrain meaning.

STIMULI: In order to maximize systematic production of nonliteral language, I have focussed upon the production and usage of figurative and literal language in the referential description of nongeometric, abstract figures taken from the game of Tangram, described above (Elffers, 1976) (as per Clark & Wilkes-Gibbs, 1986). A total of 20 figures, subdivided into two sets of 10 figures each (sets A and B), were used. Twelve figures were the same stimuli used in studies by others (e.g., Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989) and an additional eight were likewise selected for their differing geometric properties and varying degree of describability, in order to introduce a need for great detail to distinguish among the stimuli (see Appendix A).

SUBJECTS: Subjects for the production tasks were 40 females, aged 18-25, native English-speaking volunteers. Demographic information was obtained for purposes of description of the sample. Two cognitive tests were administered to each subject as baseline measures of language competence. The Boston Naming Test (BNT) was used as a measure of confrontation naming ability, and the WAIS-R Vocabulary subtest as a

measure of overall level of vocabulary. Several subjects expressed familiarity with one or another of these tests. Demographic information and test scores are summarized in Appendix B.

Experiment 1A: Spontaneous Description

The purpose of this study was to explore the written and oral production of language in the description of abstract figures. Subjects were asked to describe geometric figures on cards in the manner they felt would best enable themselves or others (peers) to later reidentify the cards. Audience condition was alternated between subjects, such that half got "self" first, and half got "other" first. Likewise, half the subjects received the set designated for oral production first, and half received the written. The stimulus sets (A and B) were alternated for written and oral tasks, such that each subject described each set for both audience conditions, but the order in which subjects participated in the two modalities was counterbalanced. Thus, half the subjects received the first audience condition for written production first, followed by the same audience condition for oral production, then the second audience condition; and the other half received the first audience condition for oral production first, followed by that of written. Each subject thus received each set of stimuli twice: once for self, in either written or oral response style, followed by the second set in the other modality, then the first set was repeated for the second audience condition, followed by the second set in the other modality. This allowed for direct comparison of descriptions between audience conditions.

Language production from the oral task was tape-recorded, and the subject was able to pause at will. Written and oral output were transcribed and analyzed for amount of

language produced, as determined by phrase count rules (see Appendix C), and content of language, as determined by figurative and literal phrase count rules (Appendix D). Phrases were divided based on grammatical parts of speech, as that was felt to be more reflective of production style differences than a word count. For the figurative/literal count, in which the distinction between the two is strongly based on context, the actual content of the sentence needed to be considered by the rater. The phrases were divided according to grammatical parts of speech and then subdivided and counted according to units of meaning, so that elaboration of a description and creation of an image was not lost.

Phrase counts were obtained for each individual description. To determine these counts, each description was scored twice, at different times, for accuracy, according to the rules. Results were analyzed as the combined set of 10 stimuli for each condition. Reliability was assessed by a trained independent rater.

An example of a subject's responses, and the scoring techniques employed, can be found in Appendix E.

For the qualitative analysis, a study of the *content* rather than grammatical role of the descriptions, descriptions were looked at in terms of whether their perspective consisted of **segmental** (unlinked but juxtaposed list of shapes or features) or **holistic** or analogical (object conceived of as a whole) properties. To study the complexity of the descriptions, the holistic perspective was further divided into "unitary", i.e., descriptions consisting of a single concept (e.g., ballerina) and more complex, multiple-concepts "multinary" (person with leg sticking out the back) subgroups (Clark & Wilkes-Gibbs,

1986). Each description was then reduced to its primary subject noun and the main subjects of the descriptions were analyzed as to their location in the categorical hierarchy, i.e., whether they were **superordinate** (e.g., person), **subordinate** (or basic) (e.g., woman) or **exemplars** of the category (e.g., ballerina). For example, "ballerina" (category exemplar) was considered more specific than the more generic "woman" (subordinate), and still more than "person" (superordinate).

To determine whether and to what degree the independent measures of vocabulary and naming might function as predictors of variance in the relationship between the independent and dependent variables, a correlation was performed.

Experiment 1B: Imposed literal production

The imposed literal task was administered after the spontaneous description tasks, because it required an explanation of literal and figurative language, and thus might have guided or structured the subjects' later responses. Upon completing the "self" and "other" descriptions, subjects were explained the differences between literal and figurative language. They were then asked to describe a subset of five figures (1-5 or 6-10) using only literal language, i.e., focusing upon the components parts of the figure and not to create a holistic, analogic reference i.e., without comparing it to another entity. All subjects performed the written task, and half the subjects were additionally asked to perform the oral task (stimuli 1-5).

Language production was analyzed using a one-way repeated measures analysis of variance in terms of the dependent variables, namely, the number of phrases (with the phrase count of the corresponding 5 stimuli descriptions from Experiment 1A) and the

number of literal and figurative phrases and the effects upon them by the independent variables of audience condition ("self" and "other", "imposed literal") and modality ("written" and "oral"). The qualitative style of the language produced under each audience condition and in each modality was measured using Chi-square (X^2) (3 noun categories: superordinate, subordinate or distinguisher; and 3 measures of the complexity of the description rendered: unitary, multinary and segmental).

Experiment 2: Identification

This study used the descriptions produced in the written production tasks of Experiments 1A and 1B to study the effectiveness of the different production conditions by having subjects identify the figures. Descriptions generated in the production task served as the independent variables (3 conditions, "self", "other", "imposed literal"), and number of correct responses was the dependent measure.

Subjects participating in the identification task were a second group of 26 female college students, aged 18-25, native English-speaking volunteers, i.e., peers of those producing the descriptions. No additional demographic information was obtained.

Methods: Each subject was presented with written instructions and a card of all ten figures from one of the two sets of stimuli numbered from 1 to 10. Each subject was given 25 randomly selected descriptions, 10 from each audience condition. Presentation order was alternated between subjects. Subjects were asked to identify the picture best fitting the description, with the added information that some figures may be described several times, and others not at all. Following these 20 sentences was a third set of 5 literal descriptions randomly chosen from the literal descriptions generated by the first

group. Subjects were asked to assign a rating (confident, possible, guess) for their level of certainty for the figure they selected. There were no time limits placed on the subjects.

SUMMARY OF PREDICTIONS & QUESTIONS

Experiments 1A & 1B were designed to study spontaneous language production in a descriptive task varying audience condition and response modality. It was predicted that in spontaneous description of complex Tangram figures, figurative/analogical language, as measured in this study, would be used more often than would literal. By expanding upon that general phenomenon, which has been documented in the literature (e.g., Clark & Wilkes-Gibbs, 1986), it was predicted that figurative language would predominate in both "self" and "other" audience conditions but the amount of figurative language would vary according to audience with "self" greater than "other", and conversely, the number of directional/literal components would be more prevalent in "other" than "self". In another measurement of audience influence, it was predicted that audience condition would affect amount and style of language produced. In terms of amount, it was predicted that the length of the descriptions (measured in number of phrases) would vary according to audience condition, with the "imposed literal" condition generating the greatest number of phrases, followed by "other", and then "self". In terms of type of language, it was predicted that descriptions for "self" would consist of more category-specific words than they would be for "other". In addition, descriptions for "other" would be more elaborated (*multinary/segmental*) than the "self" descriptions (*unitary*) (Clark & Wilkes-Gibbs, 1986).

With a focus upon the modality of production, it was hypothesized that amount

and content of output would differ between written and oral modalities. Most generally, written language would be more concise than oral (smaller number of phrases), but the patterns predicted above should be true within each modality.

In an effort to learn more about the relationship of other language functions such as naming and vocabulary, it was proposed that the relationship of WAIS-R Vocabulary score and BNT (Boston Naming Test) score to phrase count and style of production be explored. It was predicted that subjects with lower spontaneous naming score would use more language in describing a figure (i.e., larger number of phrases generated), while those who demonstrated strength in naming (high BNT) would produce more concise (shorter) descriptions and use more exemplars than superordinate category groups. Likewise, subjects with overall high vocabulary scores would produce shorter descriptions than those with low vocabulary scores.

Experiment 2, designed to study the effectiveness of the descriptions generated in Experiments 1A and 1B, predicted that audience condition would affect degree of identifiability of descriptions. Specifically, it was predicted that descriptions produced in the "other" audience condition, since they were tailored for an audience, would be more easily identified than in the "self" condition, while "imposed literal" would be least often correctly identified.

RESULTS

RELIABILITY: In order to determine whether the Phrase Count rules and Figurative and Literal Count rules could be used reliably by others to obtain the same results, descriptions by 4 of the subjects were given to another rater for scoring by the same rules. Subjects' descriptions were ranked according to total length of descriptions, and from these, samples were selected as a representative subset of the range of responses: the extreme at each end was bypassed and the second shortest and longest, and two in the middle of the range were included.

Although paired t-tests between each stimulus might appear to be the method of choice by which to compare these scoring results, it was found that due to very small mean differences between the raters, (actually indicating high reliability between the raters!) even a slight deviation between them gave rise to a significant difference in scoring, despite (or because of) the high degree of similarity in rating that had in fact caused this phenomenon. Nevertheless, figurative count difference was nonsignificant ($t(3) = 2.27$, ns, mean difference=14.0, SD=12.3), as was literal count ($t(3)=0.47$, ns, mean difference=4.0, SD=16.9). Phrase count difference ($t(3)=3.38$, $p<.05$) however, did reach significance, but the mean difference between the two raters was very small (mean difference=3.75, SD=2.2).

Using correlation to represent the interrater reliability, we obtained two r's for each set of rules, the average correlation of the individual stimuli, and also the correlation of the group of (20) stimuli by subject (4). For phrase count, the mean correlation of the set (based on 80 stimuli) was 100% (99.99) and the averaged correlation of individual

stimuli between raters was 99.6% (Range 99.4 to 100%). Thus, we see that the two raters in fact differed only very slightly in their rating of phrase count.

For figurative count, the mean correlation of the set was 98.5%, and the averaged correlation of the stimuli was 92.8% (Range 62.5 to 100%). For literal count reliability , the mean correlation of the set was 96.1% and averaged correlation of the stimuli was 90.1% (Range 70.1 to 100%).

These results indicate that the scoring criteria can reliably be used by different raters to achieve similar results, although it was seen that raters demonstrated more difficulty in the more subjective judgments of figurative and literal distinctions than they did in the more objective measurements of phrase count. Despite the significant difference seen in the more objective phrase count, the interrater correlation was higher than that of the more subjective ratings, which overall were judged more variably by the raters. In fact, it was likely this overall variability between raters that allowed the nonsignificant result. This relative difficulty with figurative/literal decisions is congruent with the reliability studies of others (e.g., Fussell, 1989), and is perhaps due to subjective interpretation of the rules either because of the rules' high degree of dependence on context, or perhaps due to lack of clarity within the rules.

EXPERIMENTS 1A & 1B: **LANGUAGE PRODUCTION:** These experiments were designed to study the differences in language production according to audience ("self" vs. "other") and modality (written vs. spoken). As described above, each subject was asked to describe sets of stimuli under differing audience conditions. Comparisons were made between audience conditions and modality of production, and the results

described below are divided into sections according to the phenomenon being studied.

The data from all subjects were included in the analyses. However, several descriptions were lost due to improper use of the tape recorder and illegibility of handwriting, so the sample size was not constant for all the analyses.

Prediction: Audience condition will affect amount of language produced.

- *Length of description (# phrases) will vary according to audience condition, with descriptions for "other" being longer than those for "self".*
- *Written language is more concise than oral for both audience conditions, but the pattern of audience differences will be the same .*

Amount of Language: Table 1A displays means, standard deviations and N for phrase count in "self" and "other" conditions for each modality. Mean written phrase counts for each individual stimulus ranged from a low of 4.2 to a high of 6.1 (Mean = 5.1) for the "self" condition and a low of 11.0 to 14.7 (Mean=12.6) for the "other" condition. The ranges are fairly small, possibly due to ease of describability of each stimulus within each condition, but it is important to note that there is no overlap between "self" and "other". For oral production, mean phrase counts ranged from 8.9 to 11.9 (Mean = 10.7) for "self" and from 20.6 to 25 (ave.=22.8) for "other", with a high degree of between-subject variability, but again no overlap between "self" and "other". For the following analyses, the individual stimuli were combined, and analyzed as a set, obtaining one phrase count score for each subject per audience condition.

Written production: Three production samples were obtained from each subject, description for oneself ("self"), for a peer ("other"), and lastly, an imposed literal

condition. Because subjects were administered only half the stimuli for the "imposed literal" condition (either the first five, or the latter five), each stimulus subset (1-5 & 6-10) was analyzed separately.

A mixed model ANOVA was used to ascertain whether effects as a result of stimulus set (A or B) or presentation order were present. This resulted in a 4 (presentation order) x 2 (stimulus set) x 3 (production condition) design. For both subsets of written data, neither set nor presentation order proved significant as a main effect or interaction. This being the case, we felt justified in analyzing the results across these conditions, and thus increasing the power of the analyses.

Because the stimuli were the same across conditions, a repeated measures ANOVA was then used to compare the mean phrase count of the three production conditions in the subsets. The between condition effect of audience ("self", "other", "literal") was found to be highly significant for both stimulus subsets in support of the initial prediction. Set 1-5: $\underline{F}$ (2,34)=13.87, Mse=667.9, $p<.001$. Set 6-10: $\underline{F}$ (2,26)=30.87, Mse=929.29, $p<.001$. Means for the subset of figures 1-5 are shown below in Table 1B and an example of a sentence in each audience condition is presented in Appendix E. The large standard deviation may be attributed to variability of individual styles of the subjects. Post hoc analysis using a series of paired t-tests adjusted with Bonferroni correction for multiple comparison was done to determine between which groups the effects lay. Results indicated a significant difference between "self" and "other" conditions t (39)=7.43, $p<.001$ and between "self" and "imposed literal", t (17)=4.43, $p<.001$. For the "other" vs. "imposed literal", t (18)=2.20, $p=.042$, which was

considered not significantly different using the Bonferroni correction (This conservative correction assigned significance to those relationships only with $p < .017$).

Another repeated measures ANOVA was performed on the complete, combined set (ten stimuli) of written data for the "self" and "other" conditions. Again, as Table 1A displays, results showed that subjects producing descriptions for another individual used significantly more language than did those producing descriptions intended simply for themselves ($F(1,39)=141.2$, $Mse=1785.36$, $p<.001$.)

Table 1A

Phrase Count for Written and Oral Modalities, by Audience (full set)

	Written		Oral	
	M (SD)	N	M (SD)	N
Self	50.8 (34.6)	40	107.1 (83.6)	38
Other	126.0 (70.8)	40	228.1 (133.5)	40

Table 1B

Phrase Count for Written and Oral Modalities, by Condition (stimuli 1-5)

	Written			Oral		
	M	SD	N	M	SD	N
Self	26.5	17.2	40	53.4	40.4	38
Other	63.6	36.6	40	113.6	74.9	40
Literal	72.5	45.3	18	185.5	83.8	19

Oral production: As with the written production tasks, subjects were asked to perform all tasks, with the presentation order alternated among audience condition ("self" and "other"), modality (written vs. oral), and stimulus set (A and B). However, the imposed literal task that followed the spontaneous conditions was administered to only half the subjects, using one subset (of five) stimuli. Again, a mixed model ANOVA was employed to ascertain possible main effects of stimulus set (A or B) or presentation order. This resulted in a 4 (presentation order) x 2 (stimulus set) x 3 (production condition) design.

Because neither set nor presentation order proved significant in the oral production task in terms of amount of language produced, we combined the data sets across these conditions. A repeated measures ANOVA performed on the three independent audience measures ("self", "other", "imposed literal") showed a strong main effect of audience condition on amount of language produced that supported the initial prediction: $\underline{F}$ (2,34)=26.59, Mse=3107.35, p<.001. Again, post hoc analysis using a series of paired t-tests adjusted for multiple comparisons using a Bonferroni correction was performed. (This conservative correction assigned significance to those relationships only with p< .017.) Results indicated a significant difference among all production conditions, "self" and "other" conditions t (38)=7.39, p<.001 and between "self" and "imposed literal", t (18)=6.75, p<.001. "Other" vs. "imposed literal" t (19)=3.55, p<.01 (p=.002).

However, because only half the stimuli (1-5) could be used in that analysis, a second MANOVA was performed comparing the more extensive stimulus sets available

for the two major audience conditions. Again, there was no effect of either stimulus set nor of presentation order was seen, and audience condition was highly significant (Mean "self"=107.1, SD=83.6; Mean "other"=228.1, SD=133.5): $F(1, 37)=70.01$, $Mse=3421.94$, $p<.001$. Again, the difference between audience condition is striking. A typical example is "dancing eagle" for "self" and "a person with with wings standing on one leg with the other leg outstretched behind him," for "other" (Subject 104, set B #3).

Summary: The above results demonstrate several phenomena consistent with the hypothesis and predictions: First, as has been documented in the literature (although not generally through parallel tasks) oral language is substantially longer than written language. Secondly, people appear to tailor their descriptions for their expected audience, using approximately twice as much language when describing a figure for someone other than themselves, in both written and oral modalities. Lastly, exclusively literal descriptions are more wordy than are the primarily analogical; this shall be pursued in greater detail below.

Discussion: Imposed Literal Description: In the final description task, in which subjects were explained the differences between analogical and literal, and then asked to describe a subset of the figures, limiting their responses to literal terms, it was observed that the number of phrases used was larger than those produced in either of the spontaneous audience conditions of "self" and "other". This result can be looked at in different ways: firstly, that literal descriptions engender greater language output because it is more difficult to create a description of this type without use of the analogical, or, alternatively, that analogical language is simply more concise and efficient than literal,

and the issue of difficulty need not be introduced.

Another alternative is that an enforced production condition, *any* enforced condition, is more difficult for a describer than is their method of choice, so more language is used while the producer of the descriptions struggles for clarity in an unpreferred modality. Were this the case, had a subject spontaneously produced literal responses for either of the audience conditions, one would expect that these descriptions would be approximately equal in length to those of the analogical responses produced spontaneously, and furthermore, an enforced analogical description would be longer than a spontaneous one, and equivalent to that of an enforced literal response.

Because it was clearly demonstrated above that figurative language was the predominant means of spontaneous description overall, it would appear that while both of the above possibilities may have elements of truth, the former may lie closer to the explanation for the results obtained herein. While many, if not most, of the descriptions produced in the "self" and "other" audience conditions consisted of both analogical and literal components in varying amounts, few (too few to allow analysis) spontaneous responses were wholly literal, without any analogical components, while the reverse was more often true, especially in the "self" condition. And because one assumes that a subject selects her spontaneous response style for clarity and convenience, the fact that analogical responses were chosen a majority of the time supports the claim that analogy is more expressive as well as more concise than literal, concrete descriptions, and is the preferred means of description of Tangram figures. Given this preference of analogical over literal, it would be difficult, at least using these stimuli, to elicit enough solely literal

descriptions to test whether an enforced figurative task would be as wordy as the enforced literal. Furthermore, the preference to the analogical over the literal is observed elsewhere in the literature (e.g., Clark & Wilkes- Gibbs, 1986; Fussell and Krauss, 1989), leading one to infer that analogical description is the method of choice for effective description, and not simply an effect limited to these stimuli in particular.

Prediction: Figurative language is the preferred means of description of complex Tangram figures.

- *In spontaneous description, figurative language is more common than literal.*
- *Figurative language will vary according to audience with "self" greater than "other".*

Written production: To test the prediction that figurative language is the preferred means of expression in spontaneous written description, paired t-tests were done comparing the total amount of figurative and literal phrases produced by each subject, for the combined audience conditions, i.e., "self" + "other". As predicted, figurative phrases (Mean figurative (self+other)=126.2, SD=54.6; mean literal (self+other) =49.4, SD=52.0) predominated, t (39)=8.46, $p<.001$, thus supporting this general hypothesis.

Effect of audience: The next analysis was performed in order to examine specifically whether the audience conditions, namely, describing for oneself or a peer, affected the type of language used. Because the "other" condition had engendered significantly more language overall than did the "self" condition (refer to Table 1A) it was necessary to use proportions in order to accurately compare *relative* amounts of figurative and literal language across audience condition. Data analyses on proportion data were performed

after arcsine transformation to ensure homogeneity of the variances.

As shown in Table 2, the mean percentage of figurative language generated in the "self" condition was 84.9%, and for "other", 73.0%. Paired sample t-tests for comparisons between "self" and "other" audience conditions for each subject were performed on the arcsine transformation of the percentage of figurative and literal ratings, for the relative amounts of figurative and literal language produced: Figurative t (39)=10.73, $p<.001$. Literal t (39)= 4.52, $p<.001$. This result supports the hypothesis that for themselves, people used more figurative language than they did for "other".

Table 2

Mean and Percentage Figurative Phrase Count, by Audience

	Written		Oral	
	Mfig /M lit	% Figurative	Mfig /M lit	% Figurative
Self	43.3 /9.0	84.9	71.4 /23.6	79.3
Other	82.8 /36.6	73.0	123.3 /81.1	69.3

Oral production: Like written production, oral production demonstrated the same tendency to overall greater figurative language usage as determined by paired t-tests of the total figurative and literal phrase counts (Mean figurative=191.8, SD=102.8; Mean literal=103.8, SD=108.2), t (39)=4.63, $p<.001$. The large standard deviations can be attributed to individual subject stylistic differences.

Effect of audience: Again, as in the written modality, because the audience condition "other" generated overall more language than did the "self" condition, an examination of

the comparison of the amounts of figurative and literal language production between audience conditions was performed using the percentage ratio of the figurative and literal language samples to the total language production for that audience condition.

Paired sample t- tests were performed on the arcsine transformations of the percentages of figurative and literal language for comparisons between "self" and "other" audience conditions, of the relative amounts of figurative and literal language produced: Figurative t (37)=3.92, $p<.001$, Literal t (37)= 3.42, $p<.01$. That the "self" condition produced more figurative language than did the "other" condition (See Table 2. Mean percentage of figurative for "self" was 79.3%, and for "other" was 69.3%) supports the prediction that figurative/analogical language is more frequently used in descriptions for oneself than in descriptions for another. Alternatively, the converse can be said: that people chose to use more literal language for "other" than for "self".

Summary: The above results support the predictions for both oral and written modalities in this non-direct communication task, and expand upon what had been previously documented in the literature in a dyadic, oral communication task (e.g., Clark & Wilkes-Gibbs, 1986). In descriptions of these Tangram figures, subjects preferred to use figurative over literal language for both audience conditions. Furthermore, subjects describing for "self" used relatively more analogical language than for the "other" condition, in which they used a greater percentage of literal language.

Prediction: Amount and content of output will differ between Written versus Oral modalities.

- *Written language is more concise than oral language.*

Area of exploration: Do written and spoken language vary in terms of relative amounts of language used, and in terms of the types of language used in description?

As has been mentioned above and shown in Tables 1A and 1B, mean phrase count differed significantly between written and oral production conditions for all audience conditions, with the number of phrases in oral production appearing to exceed that of written. This finding was established empirically through paired t-tests, which were performed between total phrase counts for each modality (oral and written) for each audience condition ("self" and "other"), "self": $t(37) = 5.70$, $p < .001$ and "other": $t(39) = 6.79$ $p < .001$. This result supports the claim that writing is a more concise means of communication than is spoken language.

Because of this overall difference in phrase count between production modalities, the absolute amounts of "other" would always exceed that of "self", and so it was necessary to use proportions/ percentage ratio in order to compare relative amounts of figurative and literal phrase counts, as shown in Table 2 (Mean percent figurative "self" for written production = 84.9%, and for oral = 79.3%. Mean percentage figurative "other" for written production = 73%, and for oral = 69.3%). Thus, to compare written versus oral production styles above and beyond simply by absolute quantity, which we have seen differs vastly, proportions of figurative and literal language for "self" and "other" audience conditions were generated. These percentages were then transformed using an

arcsine transformation and then compared using paired t- tests. "Self" (figurative: oral vs. written) t (37)=14.58, $p<.001$, "self" (literal: oral vs. written) t (37)= 3.08, $p<.01$ and "other" (figurative: oral vs. written) t (39)=-3.40, $p<.01$, "other" (literal: oral vs. written) t (39)=3.54, $p<.01$. These results indicate that written and oral language production differ in the amounts of figurative and literal language used over and above the more obvious (and possibly obscuring) difference of overall quantity of production.

Summary: In comparing written and oral language production by having subjects perform parallel tasks, we see that differences are present between the modalities, even after the most salient one, amount of language produced, has been removed. Studying the other aspects using proportions in order to control for the greater output of oral production, differences remain: the figurative to literal ratio is greater in written language than in spoken for both audience conditions (See Table 2, "percentage of figurative language"). This appears to be due to the increase in usage of literal components that manifests in spoken language, possibly in the form of reference to the external world as much as added geometric elements, since both were considered elements of non-analogical language. Additional differences, and similarities, observed between written and oral modalities are in terms of qualitative analysis, described below.

QUALITATIVE ANALYSIS: Qualitative analysis of the descriptions was conducted at several levels. These analyses focused upon the content of the language produced rather than the quantity of a certain type (figurative or literal) of language.

Prediction: Audience condition will affect the style of language used.

***1) Specificity of response*:**

- *Descriptions for "self" will be more specific than for "other" (i.e., more category exemplars vs. superordinate).*

As described above, the main subjects of the descriptions were divided into groups according to their level of category membership, i.e., into superordinate, subordinate or exemplars of the category. It was predicted that for "other", where common ground needs to be established between the communicators, more superordinate or generic category terms would be used, while for "self", where common ground is maximal, more distinguisher, highly specific nouns could, and would, be used. Subordinate categorization was considered an intermediate response.

The categorization values of superordinate, subordinate and segmental were all intrinsically linked since each stimulus could be assigned to only one of these groups. Therefore, a series of 2x3 Chi-square analyses were performed to determine the relationship between the audience condition ("self", "other") by each of the categorization groups for each of the stimulus sets (A, B).

Written data: Chi-squares were performed on each set of stimulus data in order to determine whether the sets A & B were equivalent. As shown in Table 3, it was found that X^2 (df=2, n=800)=1.69 was nonsignificant, and thus the sets were combined and analyzed as a single set. It was observed that 31% of the "self" responses were superordinate and 29% were distinguishers, compared with 59% and 15% respectively of the "other" responses (see Tables 3). To see whether these numbers differed statistically,

a 2x3 chi-square determining whether audience condition had an effect on the specificity of category noun used by the subject found that X^2 (df=2, n=800)=64.04, p<.001. This indicates that as predicted, subjects describing for another person were more likely to use a more general, superordinate noun than when describing for themselves, and likewise, to use more distinguishers or specific nouns when describing for themselves.

Oral data: As shown in Tables 4A and 4B, an initial chi-square found stimulus sets A and B to differ significantly, so they were each analyzed separately. In set A, the distribution in the self condition was 40% superordinate, 32% subordinate and 24% distinguisher categories, while in the "other" condition, the distribution was 47%, 42% and 12% respectively. Percentagewise, the "other" condition appeared to have fewer distinguisher nouns. This was confirmed statistically with a chi-square, X^2 (df=2, n=400)=22.12, p<.001. Set B had a distribution of superordinate 46%, subordinate 27%, and distinguishers 27% for the self condition, and 65%, 21% and 14% for the other condition. X^2 (df=2, n=400)=16.96, p<.001. So although there appeared to be certain differences within the stimulus sets with regards to type of noun used, the effect held constant that subjects describing for "other" would use more superordinate, general descriptions than they would for "self". This was seen to be the case for both oral and written modalities. Some typical examples of the audience differences in terms of subject specificity are as follows: "monk", and "villain from Boris and Natasha" for "self" in contrast to "person" and "small person" for the same stimuli in the "other" condition.

Table 3

Noun Categorization: Written Production

Sets A &B combined

	Superordinate		Subordinate		Distinguisher	
	N	%	N	%	N	%
Self	125	31	158	40	117	29
Other	237	59	101	26	62	15

Table 4A

Noun Categorization: Oral Production

Set A

	Superordinate		Subordinate		Distinguisher	
	N	%	N	%	N	%
Self	76	40	61	32	53	28
Other	98	47	89	42	23	12

Table 4B

Noun Categorization: Oral Production

Set B

	Superordinate		Subordinate		Distinguisher	
	N	%	N	%	N	%
Self	89	46	53	27	53	27
Other	128	65	41	21	27	14

Written versus Oral Modalities: In order to determine whether the distribution of noun categorization was similar across modality, the written and oral responses were compared. Due to differences between the stimulus sets that were mentioned above, each set was analyzed separately. For set A: X^2 (df=2, n=800)=4.28, ns, and for set B: X^2 (df=2, n=791)= 8.52, $p<.05$, indicating a significant difference between the written and oral production styles. This difference was seen in the distribution of superordinate and subordinate reference (written =55%, 24%, 20%, oral= 47%, 33% and 20%), while the relative amount of distinguisher reference remained the same.

***2) Complexity of Response*:**

- *Holistic, analogical responses will be preferred over segmental,* i.e, Production of Holistic (unitary + multinary) > Segmental.
- *Descriptions for "self" will be less elaborated upon than for "other" (more unitary responses than segmental,* i.e., Unitary (self) > Unitary (other); Segmental (other) > Segmental (self). Multinary responses were considered to be more generic, neither general nor specific, and not audience-dependent.

This portion of the analysis was conducted only on the written production data set; since the oral sample had a high phrase count, most of the descriptions were highly elaborated, and thus there existed little likelihood of unitary descriptions being produced, irrespective of audience condition.

As described above, descriptions were analyzed according to their complexity: holistic (unitary and multinary) and segmental. A series of 2x3 chi-squares were performed to determine the relationship between the audience condition (self, other) by

each of the categorization groups (holistic unitary, holistic multinary, segmental) for each of the stimulus sets separately (A, B) (see Tables 5A and 5B).

As shown in Tables 5A and 5B, stimulus sets A and B were found to be nonequivalent (X^2(df=2, n=800)=9, $p<.05$), and therefore were analyzed separately. Overall, holistic (unitary and multinary combined) responses predominated over segmental responses. This is illustrated in Table 6. (In set A, holistic responses constituted 91% for "self"; 85% for "other"; in set B, they made up 96.5% for "self"; 86.5% for "other".) In all holistic responses, in sets A and B together, there was only a single incidence of a unitary response in the "other" condition, as compared to 23 and 43 occurrences in the "self" condition in sets A and B, respectively. (The small cell size then, would be due to the strong effect of the condition.) Likewise, segmental responses appeared in only 19 and 7 instances for the "self" condition, but 30 and 28 times in the "other" condition. Both sets A and B demonstrated significant effect of audience, with X^2(df=2, n=400)=25.91 and 54.06, respectively, $p<.001$. These results seem to emphasize that while analogical responses predominated in both audience conditions, when subjects were describing for others, they rarely used simple, unelaborated responses as they did for themselves.

Table 5A

Complexity of Response, by Audience

Set A

	Unitary		Multinary		Segmental	
	N	%	N	%	N	%
Self	23	11.5	158	79	19	9.5
Other	0	0	170	85	30	15

Table 5B

Complexity of Response, by Audience

Set B

	Unitary		Multinary		Segmental	
	N	%	N	%	N	%
Self	43	21.5	150	75	7	3.5
Other	1	0.5	171	85.5	28	14

Table 6

Holistic and Segmental Responses, by Audience

	Holistic				Segmental			
	Set A		Set B		Set A		Set B	
	N	%	N	%	N	%	N	%
Self	181	91	193	96.5	19	9	7	3.5
Other	170	85	172	86.5	30	15	28	14

Summary: The above contingency studies focus upon the qualitative differences of the response style according to intended audience, and when feasible, modality. Results indicate strongly that subjects use more specific, less elaborated descriptions when the description is intended for their own use, but when it is intended for another, the subjects' responses become more highly elaborated, often less analogic, and more generic, i.e., in a more basic or general category. Many subjects used intermediate responses for both audience conditions, but the balance changed on either side of the intermediate, noncommittal responses as a result of condition.

3) Naming, Vocabulary and Language: **Relationship between Vocabulary & Naming to Amount and Style of Output**

Naming and vocabulary scores were obtained from each subject as a baseline measure of linguistic competence, as shown for the subject group in Appendix B. It had been hypothesized that subjects scoring lower on the naming task would use more language (greater number of phrases) in describing a figure. Those with strength in naming (high BNT) and those with high vocabulary scores would produce more concise (shorter) descriptions than those with low naming and low vocabulary scores. Likewise, it was predicted that high BNT and high Vocabulary subjects would use more exemplars than superordinate category groups and more consolidated descriptions than would low BNT and Vocabulary subjects.

In the results of a series of correlation studies, there appeared to be no obvious systematic relationship between the variable of vocabulary, as measured through WAIS-R Vocabulary, and language production measures, namely that of phrase count, i.e., length

of descriptions or amount of figurative and literal phrases, nor between that of naming ability as measured by the BNT with these language production measures. Scatter plots for these variables were made in order to explore whether any nonlinear relationships were present; these too showed no obvious relationship among the variables.

Because of our desire in this study to maintain a homogeneous subject sample, so as not to confound the language production data with main effects from education, or sex or age, etc., it remains possible that the above hypotheses are in fact true, if given the opportunity to manifest. If this is the case, this effect was masked in our sample, using these stimuli and these scales.

Discussion: Vocabulary & Naming: A relationship between a high/low vocabulary level and likewise, a high/low naming level to the amount and style of language output in the descriptions of the Tangrams had been predicted. As mentioned above, no obvious systematic relationship, linear or not, was observed. This lack of relationship observed between naming, vocabulary and amount and style of output might be attributed to several possibilities: 1) The effect does not exist at all. 2) The effect does not exist in this sample. The former is possible, but unlikely, given that naming ability and vocabulary size are prerequisites to, and necessarily linked to communication. If the latter is the case, several possible reasons may exist. It may be due to the small sample size, although the size of the sample was large enough to obtain a main effect of other conditions under study. More likely, it is due to homogeneity of the sample, selected to be so in order to minimize or eliminate main effects of just these kinds of demographic differences, so as to focus upon the manipulation of audience. With such a sample, the range of naming

abilities may not be great enough to observe a difference (see Appendix B for test score ranges). Alternatively, it may be due to excess ease of describability of the stimuli. Perhaps with other stimuli, vocabulary differences would become more apparent, as would naming in a task that consisted of more obscure referents. Lastly, this result may be due to insufficient measures of differences in vocabulary level, again linked to homogeneity of the sample, or naming level, since the BNT is generally used in pathologic populations and little variance was seen in the scores.

It is likely, or at least possible, that given a more diverse population which ranged in age and/or education, that differences in production would vary, thereby giving us information about the interrelationship of these variable with language production. In a large population, moreover, it might be possible even to glean the degree of the interrelationship, and the interaction of high vocabulary/low naming and the converse, and how these strengths and weaknesses might influence language production. It must be remembered though, that these effects might prove so strong that they would make other comparisons, such as the ones we were primarily interested in in this study, more difficult or impossible to evaluate.

***EXPERIMENT 2*: COMPREHENSION/IDENTIFICATION**: This study was performed in an attempt to discover whether the descriptions produced under the differing audience conditions ("self", "other", "imposed literal") resulted in a third party's differential ability to identify the pictures described.

Prediction: Audience condition will affect degree of identifiability of descriptions, i.e., descriptions produced for "other" will be more easily identified than for "self", and

descriptions produced in the imposed literal condition will be least easily identified.

Subjects were presented with 25 descriptions to identify: ten descriptions intended for "self", ten for "other" (order alternated), and lastly, five literal descriptions, as described in an earlier section. Because of the inequality of the size of the stimulus sets, calculations were performed using percentage correct, but below are discussed as integers (i.e., sums of literal descriptions were doubled).

Mean correctly identified descriptions produced in the "self" condition was 6.3 (SD=1.3), for "other" 7.0 (SD=2), while for imposed literal, the mean was 4.6 (SD=2.5). A 2 (presentation order "self" first or "other" first) x 3 (description types: self, other, literal) mixed model ANOVA was performed. The analysis showed there to be no main effect of presentation order, but a significant effect of audience F (2,48)=9.31, Mse=3.54, $p<.001$. Post hoc analyses were done using paired t-tests, with Bonferroni correction for multiple comparisons (in which significance was reached at $p<.017$), to determine between which conditions lay significant differences.

It was found that there was no significant difference between the number correctly identified for the "self" and "other" conditions (t (25)= 1.48, ns)), but the descriptions produced in the imposed literal condition were identified correctly significantly less often than were the others (self/literal t (25)= -2.84, $p< .01$ ($p=.009$); other/literal t (25)= -4.46, $p<.001$. This result runs partially counter to the initial prediction, that descriptions generated specifically for another would be more readily identified than those reserved for oneself which had been observed in earlier studies using a similar task (e.g., Fussell & Krauss, 1989a,b), but supports the prediction that literal would be the most difficult to

identify.

There was found to be no relationship between the number correctly identified by each subject and the level of certainty of the response expressed by each subject.

Summary: These results indicate that a peer group of the initial subject pool could identify descriptions produced under the "self" and "other" audience conditions equally well, but had more difficulty identifying those descriptions produced under the imposed literal condition.

Discussion: Identification Task: There were a great number of differences noted between audience conditions in terms of amount and style of language production, in that subjects, when describing for the "other" condition, tailored their descriptions to the audience by using more language, and more literal phrases in order to feel certain that a peer would be able to use their description effectively. There were also substantial differences seen in the type of language, e.g., in terms of the specificity and complexity of the primary subject nouns. Interestingly, though, despite the producers' efforts to create a distinction by audience, there were no significant differences in terms of the number correctly identified in the identification task by audience. This brings to light an interesting phenomenon concerning people's perceptions of others; specifically, subjects may not have an accurate sense of their listeners' abilities, and therefore they add detail, even while it is unnecessary. The possibility that the audience is "smart" might have been overlooked by the description producers, or alternatively, due to lack of audience feedback, subjects are just making certain that their descriptions will be understood by elaborating upon the descriptions and assuming only a least common denominator of

receiving skills.

It is also possible that the identifiers in this study are too similar to the producers; if they were not peers, perhaps a differential would be more apparent. While common ground is only absolute in descriptions for oneself, with peers, there would be some natural overlap of common ground that might reduce any difference between the audience conditions. In contrast, though, the subjects producing the descriptions with their substantial stylistic differences were aware that they would be judged by a peer, and formulated their descriptions accordingly- i.e., differentially by audience.

Furthermore, the subjects were not out to prevent someone from successfully identifying the descriptions for "self", so it is natural that what would be easily identifiable to them would be similarly identifiable to a peer. Had the subjects been asked to describe the figures in a code, i.e., in an idiosyncratic way only they would know, there would likely have been an audience condition difference in identification of the figures. Likewise, it is unclear what the quality of these secret personal descriptions would be.

Differences were seen in the comprehension/identification of the literal descriptions relative to the audience conditions of "self" and "other"; several possible reasons can be advanced for this. Because literal descriptions consist of knowledge that we all share, and as such are remarkably context-free, one might speculate that these descriptions would be the most accurately identified of all. However, results were more congruent with the initial prediction, namely that literal descriptions would be identified least accurately. An explanation for this poor performance in the identification of the

literal descriptions might be that attention and interest were at a minimum, or were lost; just as most subjects had difficulty producing the literal descriptions, and often became repetitive or disorganized, so might the identifiers have had similar difficulty receiving these long, sometimes convoluted messages. Finally, in terms of the lack of relationship of the confidence ratings (definite, probable, wild guess) of the subjects' responses to their actual correctness, it is possible that the lack of time constraints imposed on the subjects allowed them to expend different amounts of effort on this task, leading to a speed/accuracy tradeoff by some; or due to a more general lack of certainty in terms of distinguishing among many similar stimuli; or to an inconsistency of subject sophistication concerning subjective rating scales.

Discussion: Description Quality: The issue of quality of the descriptions is one that must be mentioned. In this study, the descriptions produced were not individually prejudged for "quality"; all descriptions were randomly included in the identification sample. Presumably, though, some descriptions were "better" than others, but how would this be judged except by the number that could be correctly identified? Even that would not be a foolproof method of rating, since raters vary in motivation and dedication just as do the producers of the samples. Even descriptions considered "poor" might be identified correctly by a highly motivated or intuitive rater. Therefore, no quality rating was done on the descriptions generated, and it was hoped that any inconsistencies in quality were controlled for somewhat by the randomized order, and by the fact that all subjects received all conditions, so the effect of a particularly diligent or lax describer would be balanced, and similarly the peer group used to identify the descriptions also received all

conditions, so the effect of a less-than-consistent subject might be absorbed.

GENERAL DISCUSSION

RESULTS SUMMARY

In this study, we have attempted to study language production by manipulating modality of communication and audience condition. Previously, most research has concentrated upon the comprehension of language, and has demonstrated a strong effect of context. The studies described herein have focussed upon the aspect of language *production* in context, and have tried to add to the research knowledge base that has until now lagged seriously behind that of its comprehension. In this dissertation, it was hypothesized that under more general conditions (audience "other") describers would utilize more literal figures of speech, which can be understood by many, while in more specific conditions ("self"), they would select from a larger range of possibilities that may be more idiosyncratic, more figurative, and overall more difficult to be understood by others, even while facilitating their own personal comprehension of the referent. Results supported these predictions, confirming that the intended audience plays a major role in determining what style of language will be produced, and that while written and spoken language do differ from each other in many ways, they reflect similar patterns of production, as shown in Tables 7 and 8.

Table 7

Production Patterns, by Condition

	Written	Spoken
Amount language	Literal>Other>Self	Literal>Other>Self
% Figurative	Self >Other	Self >Other
% Distinguisher Nouns	Self >Other	Self >Other
Number Unitary:Segmental	Self >Other	ND

Table 8

Production Patterns, by Modality

	Modality
Amount language	Oral >Written
% Figurative language	Written > Oral
% Distinguisher Nouns	Written = Oral

Additionally, parallels were noted between comprehension and production: In comprehension, context allows figurative meaning to be understood automatically; likewise we see that context determines whether and to what degree figurative language will be produced. In comprehension, the intention of the speaker is understood by the listener/reader. We see that same effect in production, in that the producer's intention, created by context, determines how her thoughts will be conveyed (and furthermore, that they are understood). Thus, we see that, in the limited scope of this dissertation, production of analogical language versus literal language could be manipulated by

specific circumstances, as was predicted.

Although the above studies demonstrated the occurrence of this linguistic phenomenon, we certainly cannot hastily generalize the results to all situations. Some of the limitations of this study, described earlier, were deliberately built-in, included specifically in order to minimize variability and control for main effects by extrinsic factors. Other limitations were in the methodology: for example, the stimuli were selected to generate relatively limited descriptions rather than free-form spontaneous language, and the language generated can be analyzed only with the caveat that all analysis is subjective and therefore open to interpretation and thus to questioning. Furthermore, it was seen that even while scoring techniques could govern fairly reliably an objective measure such as phrase count, upon presenting raters with a more subjective, context-dependent measure (such as figurative vs. literal judgments), this high interrater correlation decreases slightly. It is possibly for reasons such as these that most research on nonliteral language focusses primarily on clear-cut figurative manifestations, such as metaphor, simile, and metonymy. However, as mentioned earlier, spontaneous occurrences of innovative instances of these phenomena in the language are not sufficiently prevalent or systematic to be studied. If we assume that *all* language is figurative, the language sample is large, but since all language is figurative, we can say little concerning figurative vs. literal production. This theory of figurative language is largely a philosophical or anthropological exercise, although it is perhaps revealing in terms of the interlinked roots of language and cognition (e.g., Lakoff & Johnson, 1980).

Therefore, the very same drawbacks and limitations of this study are also its

advantages, in the sense that an effort has been made to study the components of spontaneous language (albeit controlled). We have tried in the pages above to stress the nebulousness of the boundary between figurative and literal language. That trends exist in spite of so many individual subject differences may speak in favor of the robustness of the effects.

Finally, we reiterate our contention that the preference toward the analogical described in these studies *is* in fact a demonstration of figurative language in the real world in that people use words with the intention of creating an image that enhances their listener's comprehension of their words. It is *intention* that determines what role a word or phrase will play at a particular point in time. It is the use of analogy in context, in this case, the comparison of one thing, geometric shapes, to another for the purpose of effective communication. In the sense that figurative language is creative, the language produced in these studies, is.

CLINICAL IMPLICATIONS AND DIRECTIONS

The studies described above can be considered a preliminary exploration in the area of figurative language production, and more generally, in the systematic study of language production and the effect of context on this production. While this study was limited in scope, specifically in order to focus upon context without intrusions of other variables, this limit could have obscured the effects of just these same variables which play an important role in real-life language production. Some of these factors have been mentioned above, such as sex of the subject, differing relationships among the participants, range of education of the participants, and intellectual functions such as

vocabulary and naming ability, all of which may play a role in determining word choice and communication style over and above the constraints of the situational context. While pilot data suggested that differences exist among age groups, and between sexes, these have not yet been studied systematically and it remains unclear whether and to what degree these variables impact upon production.

Although research efforts in this area have not been extensive to date, it appears to be an area of burgeoning interest as the disciplines of psychology and linguistics are ever more intersecting. Simply an awareness of the above factors' possible role in the type of language produced might lead to a greater understanding of communication, especially in the areas of dysfunction.

An example of this can be seen in the right-hemisphere damaged population, who although they may appear unimpaired when observed superficially, are known to be concrete and literal in their comprehension of language at all levels, (e.g., Brownell, 1984), insensitive to alternative interpretations of words (Brownell, et al., 1990), deficient at discerning nuances in the speaker/listener relationship (Kaplan, et al., 1990) and demonstrate an impairment in holistic/segmental perception of visuospatial stimuli, such as in their ability to synthesize a whole from its parts (Heilman & Valenstein, 1985; Lezak, 1983). It is as yet unclear whether this synthesis impairment is present in verbal descriptions of figures as well. In light of these differences in comprehension and perception, it would be of interest to see whether their linguistic production could be manipulated, i.e., whether their descriptions would vary with context, and whether the content parallels that of the normal population; and if not, how it differs. A greater

understanding of the language production styles of these patients would provide further knowledge as to whether the language deficits seen in their comprehension style extend to production as well, and thereby mirror the cognitive deficits they manifest. If this were seen to be the case, namely, if, for example, a stroke patient were found to be exceedingly literal in terms of comprehension, it would be of value to determine whether his style of production paralleled his comprehension, and if so, could conditions be altered to *elicit* certain types of responses. Were this possible, it would increase the understanding about the deficit, and open the door to better compensatory, rehabilitative efforts.

PSYCHOLINGUISTIC IMPLICATIONS

The results above generally support the hypotheses and confirm the predictions that accompany them. There is more or less clear-cut evidence that language output varies as the audience conditions change. Although perhaps never studied as systematically, this result in itself is not wholly new. Rather, the issue of importance under study herein is what the changes were, the circumstances in which these changes manifested (and did not), and whether any information from this knowledge can be extended beyond the realm of general interest. In this study, we have considered language a reflection of thought; therefore, a study of word selection and how it varies in context can and must be considered as a deliberate process that reflects differing perceptions and communicative needs.

Like art, language is a symbol system that has evolved for people to communicate their representations of experience. Our conceptions of reality have culminated in symbolic expression as a means of understanding and preserving experience, and

communicating it to others. Language is limited in the way all symbolic forms are limited -- once something has been defined, it is confined to the specific parameters of that particular modality's definition of it, while certain of its other, peripheral attributes may become obscured or lost. Yet words, however imperfectly, do still reflect our thought processes, and de facto, must be seen as more than an imprecise record of reality. Instead, language can be considered an original means of expression in which the words help create reality, since only with their existence can anything become an object of thought (Cassirer, 1946). Wheeler (1990) extends this view, to contend that language, and particularly metaphoric language, even has the ability to change reality, or rather, reality as we perceive it. We use language as a medium to construct and organize our world-- its categories, its concepts, and by so doing, our reality.

Because each symbolic form has its unique perspective, its own way of seeing an object, in a way, language does help form the way we see the world. The mind forms concepts by taking a number of objects, and comparing their similarities to arrive at concepts, classes, categories. Thus, a word comes to exist in its own right, as an objective reality, not simply as an arbitrary symbol. Over time, it has come to be that a meaning and its word appear as one and the same. The equivalence of a word and its meaning has enabled language to itself become an object of inquiry: The words chosen by a speaker indicate the perspective s/he holds in viewing the world. Thus, the study of language is a source of knowledge about both the conceptual system of the world and of the speaker; and the words he uses, which he selects over other acceptable alternatives, reflect the current external reality in conjunction with internal cognitive processes.

Elaborating upon the theory of Lakoff and Johnson, Gibbs (1994) has explored the issue of whether how we speak about our experiences is linked to the ways that we conceptualize our lives; he suggests further, that figurative language *relies* on context and real-world knowledge, and that people conceptualize experience in figurative terms, and thus metaphor both "motivates and constrains" how we think creatively. Constraints on how we speak or write are not imposed by linguistic restrictions, but rather, by how we think of experience. Thus, descriptions and phrases, in both prose and poetry, that may appear to be wholly novel, in fact are often manifestations of the possibilities extant in a conceptual metaphor already shared by many in a culture (p7).

One finding of this dissertation has supported the earlier research finding that figurative language is a more concise form of expression than was use of strictly literal terms. This result seen on its own might appear to be counter to that seen in poetry, in which language expression often appears to be flowery- or very terse- simply for esthetic reasons. The theory of Gibbs, however, suggests that while some poetry might seem to be "only" artistic, it might also be a means of referring to something that is otherwise difficult to express, and which due to this inexpressibility, requires more words to describe than does description of a concrete object. Creation of reference is, in a sense, a form of categorization: a group of features is recognized as a whole and given a name. When the features are obviously joined, as in a concrete object, the object can be named at different levels according to the situation and the specificity required; for example, several pieces of wood in a particular arrangement can be a barstool, a chair, or a piece of furniture, in ascending order of generality from distinguisher to superordinate. Because

these features are so clearly defined, they are considered to directly represent the object under description, in a manner that will be comprehended by others, and as such, each is considered a literal term for the object.

In contrast, in a situation where features are less clearly linked, more ambiguous, or less representative of an object tangible to all, the creation of reference becomes more complicated. In this case a description of the individual parts might be considered literal, but because the pieces are not necessarily or definitely joined together, a holistic description is really only an approximation, or an analogy, a form of figurative description. (Such were the stimuli in the experiment described above: the representations were of concrete objects, but only when the subjects determined the context was sufficient to link the parts. When the context was undefined, as in the "other" condition, the pieces more often remained unlinked.) It is at this stage that the holistic representation is more concise than the segmental, literal descriptions of the individual pieces, but nonetheless speakers must judge whether they will be understood, and accordingly trade off length for comprehensibility .

And when something more ambiguous is described, such as an emotion, to ensure comprehensibility, speakers will use words to create a representation that cannot necessarily be visualized. They must try to elicit in their listeners or readers an image that is not clearly symbolized in memory like a concrete object, but is rooted elsewhere in their experience. To do so, they might resort to evoking a conceptual metaphor that is already familiar, and that when elaborated upon, leads to the shared experience. Thus poetry might not differ so greatly from other forms of descriptive language- like all forms,

it strives to evoke an image; unlike other forms, it often must do so without a concrete target of representation.

The results of this dissertation lead to certain conclusions, or at least to some speculations, regarding human behavior in regard to language production, and regarding language production as a reflection of cognition, insofar as output mirrors the compilation and analysis of internal and external factors. A more general consideration, inclusive of, but extending beyond, the issue of a subject's frame of reference, is one not necessarily concerned with the responses produced in either the "self" condition or the "other" experimental condition, but rather with the differences generated between the conditions. Differences in language reflect differences in the perception of exactly what is needed for communication, such as what the common ground is, and what needs to be conveyed. Consideration of our results in this light places a focus upon the means by which different linguistic responses are produced in terms of understanding how and why a particular word is selected over another. When differing responses are generated in different contexts as systematically as demonstrated in the preceding pages, it can be safely suggested that this was more than a random occurrence. A possible model for such actively changing responses in the face of context is the Competition Model of Language (MacWhinney, 1989) , which derives in part from the principle that there is a fundamental constraint against synonymy, i.e., no two words are exactly the same, and if one is selected over another, there must be a reason. Two considerations must interact: 1) a word can mean anything, in a given context, and 2) each word is chosen for a reason. MacWhinney (1989) proposes that differing cues favoring a match between an object and

a word vie with each other, and that the candidate with the greatest cue strength wins. The idea of words competing yet cooperating within a framework is enlightening in suggesting a relationship between word meanings and semantic topography. But the notion of a fixed topography, in which words each simply occupy a niche in the semantic space, is too limiting to explain certain areas of language processing, such as novel constructions and linguistic innovations. He suggests rather, that we think of words not as just finding their places in a preformed landscape, but as also helping to shape that landscape (p220). The words compete in different situations, and through repeated exposure, the topography covered by that word becomes adapted to new competitive influences, either through repetition of well-worn pathways, or alternatively, through development of more innovative meanings that expand into new territory of the semantic map.

The conversational models in the literature which discuss cooperation between the speaker and the listener might also help explain how the landscape is shaped. Factors such as intention, and context must act as driving forces in determining what word will be used in a particular situation. For example, the words *sad, morose, glum, unhappy*, and *blue* all convey a sense of the (-happy) each with its varying nuances, degrees and connotation. For a word to be selected over another, several events must occur: First, and necessarily, it must be available in a person's lexicon, and accessible on demand. Then, it must best denote, connote, and express the intention of the speaker for that situation, at that moment in time. How this latter process takes place can perhaps be explained through uniting the principles of the competition model with the interactive elements of

the conversational model. In other words, the word selected must reflect the summation of many different processes, i.e., it must take into account the speaker's assessment of the audience's understanding, sometimes with, and sometimes without (as in the present study) direct feedback.

Models describing the assembly of meaning have ranged from the early classical theory of feature-matching, in which membership in a class is categorical, to the more fuzzy prototype model which is more true-to-life, using weights and resemblances, to the competition models (including PDP as a computational model), that describe a continual decision-making process. According to the competition model of MacWhinney (1989), and adapted by Gibbs (1995), words are dynamic, active participants in communication. Through words, people *create* meaning. The idea of a dynamic, continual decision-making process reflected in this model of language production, and specifically figurative language, is a real-life model of language that takes into account that many factors are working simultaneously, exerting influence on the output. What is lacking in that theory are the types and the degree of external influence.

It is in this area that I suggest a linking of the categorization and competition models with the conversational models, in terms of the way and the amount in which they influence word choice. The former models acknowledge many possibilities, and the latter observe what the choices were. Joining the two might allow us to manipulate in a systematic way, conditions to elicit output, to account for the power of the speaker's specific intentions, and to predict the conditions best suited to a particular mode of production. Claiming that words exist on a denotative/connotative continuum begins to

express this greater flexibility, but still fails to fully communicate their elasticity. To do so, we might consider the meanings of a word to have a normal, bell-shaped distribution. (This analogy emerged through discussions with V. Tartter.) The denotative, most standard aspects of a word are at the center, with lesser-used meanings on either end, and even further from the "core" are outliers, such as novel usages of words brought in for a special situation; these might be better known as members of another curve, but at times, find themselves overlapping another set. Certain features, such as context, move a word along its meaning curve, placing it at different spots on the curve at different times.

The results of this dissertation are consistent with this analogy to a normal curve. It was seen that different types of nouns were produced under differing audience conditions- under both conditions, subordinate nouns predominated, but the audience condition determined the precedence of the outliers, i.e., the distinguisher and the superordinate nouns. This result may have support in a more general sense as well; for example, Lakoff and Johnson (1980) proposed that all concepts are understood in terms of other concepts, and thus appear to endorse this idea that at different times, different properties are drawn upon from other areas of cognition. And likewise, Gibbs (1994) suggests, in his discussion of poetry, that properties are imported from other concepts to create new images, or new ways of looking at something familiar.

What is still lacking is a model that can more specifically describe the conditions for language production, one that combines and synthesizes features of the competition model, in which words, with all their different meanings, vie for prominence, with the elements of the conversational model, suggesting the interactions of external causes with

the internal for creating movement along the meaning distribution.

Literal and figurative meanings are not well-defined, and the boundary between them is as fuzzy as the boundaries of any category, or between any two synonyms, and multiple factors continually work to determine what *exactly* is intended and best to be utilized in any given moment. According to context, and determined by the current situation, the background of the communicators, the intentions of the speaker as to what exactly needs to be conveyed, and the ability of the recipient to grasp what this may be, the best word or phrase is selected over a multiplicity of others. Generally the selection is correct for the situation, and imparts (presumably successfully) whatever it was the speaker intended. However, at times errors may occur, errors such as paraphasias, which resemble the desired word either phonemically or semantically; or if the target word cannot be retrieved, a TOT error might result, in which all aspects of the word aside from the word itself might be available. In cases such as this, rather than simply choose from a large repertoire of lexical alternatives, a person will struggle to recover that *particular* word which best conveys his intention.

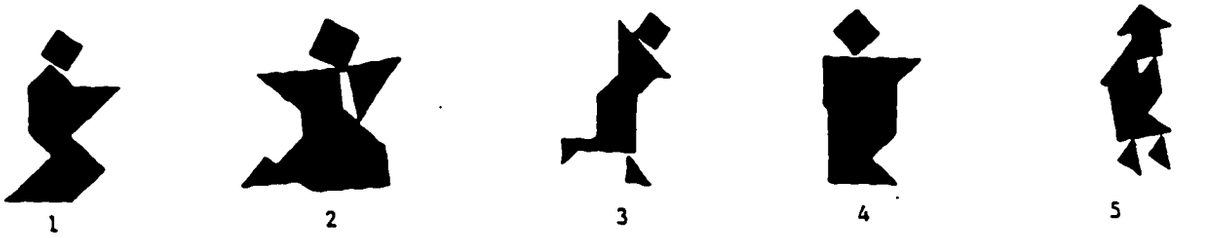
While all speakers of a language communicate fluently and effectively (in varying degrees), people speak differently from one another, using different styles and somewhat different vocabularies. Words and language style are preferentially selected over others through filters created by differences in education, abilities, contexts, needs and interactions. All these aspects must be factored into any model of language production. Results from this dissertation have provided evidence of effects of external context on language production, and although there was no clear evidence of the impact of

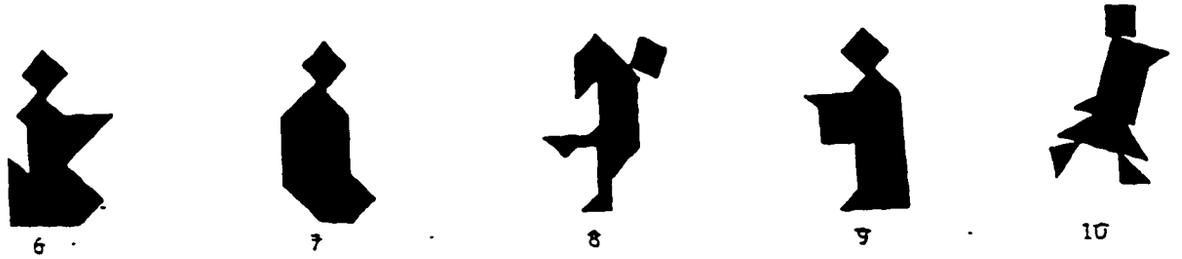
intellectual functions, that area is still open to exploration. The topic of language production and the factors that impact on it are only now beginning to be studied in depth, and have yet to be explored on a large scale, taking into account the many complicated interactions that surround, and moreover, *create* language.

APPENDIX A

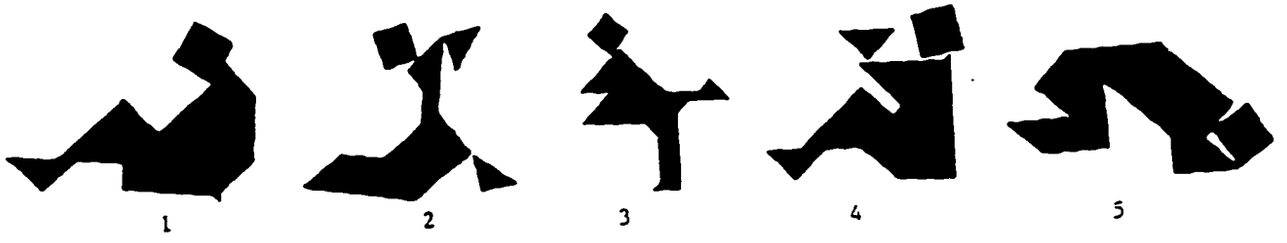
TANGRAM FIGURES

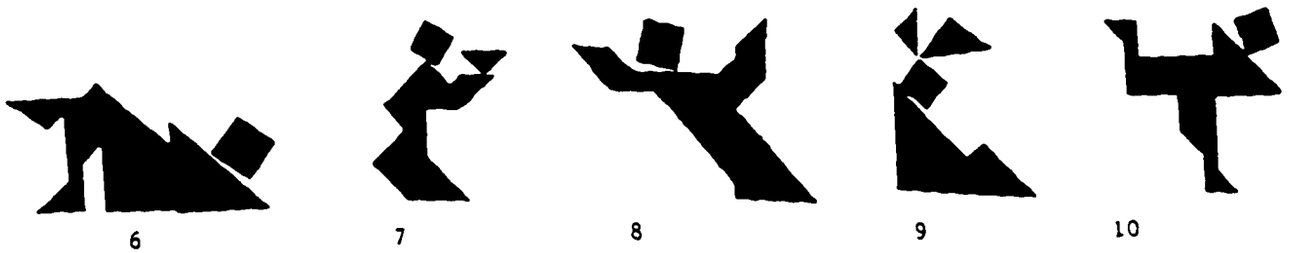
SET A





SET B





APPENDIX B

SUBJECT CHARACTERISTICS

	M (SD)	Range
Age	22.4 (2.1)	19-25
Education (years)	15.6 (1.5)	12-18
WAIS-R Vocabulary (raw)	58.5 (5.7)	48-68
BNT (raw)	54.6 (2.8)	48-60

APPENDIX C

PHRASE COUNT RULES

	Example	**phrase count**
General:		
NP= (det)(adj) N	A (tall) man	1
PP= P NP	Under the hill	1
VP= (aux) V	Is running, bowing, resembles	1
V(prt) NP	Stretching out his arm	2
V PP	Running down the hill	2
V adj	Is bent	2
Some Rule Combinations:		
N PP	Arm behind his back	2
N adv	Arms up	2
NadvPP	Arms up in the air	3
N or N	Running or walking	1
N and N	Arms and legs	2

Key:
N= Noun
V= Verb
P= Preposition
adj= adjective
adv= adverb
det= determiner (article)
prt=particle
xP=xPhrase

APPENDIX D

FIGURATIVE/LITERAL COUNT RULES

In addition to the rules, which are grammar-dependent, context must play a role in determining whether a phrase is figurative (F) or literal (L). If a word is geometric or shape related (e.g., square), or relates to the world outside of the description (e.g., to the left of the index card) it is literal. If it creates or extends an image or analogy, it is figurative.

	example	**count**
General:		
NP= (det)(adj) N	A (tall) man	1+ 1
per adj		
	A square head	1F 1L
PP= P NP	Under the hill	1F
VP= (aux) V	Is running, bowing	1F
V(prt) NP	Stretching out his arm	2F
V NP (NP required)	resembles a man	1F
V PP	Running down the hill	2F
V adj	Is bent	1F
Some Rule Combinations:		
N PP	Arm behind his back	2F
	Arm to the left	1F 1L
N adv	Arms up	1F 1L
NadvPP	Arms up in the air	2F 1L
N or N	Running or walking	1F
N and N	Arms and legs	2F

APPENDIX E

SAMPLE DATA: SCORING

Subject #119, stimulus set A, figure #1.

Self: **A guy sitting on a train reading the paper in the morning.**

	NP (det N)	VP-V PP	VP- V NP	PP
Phrase	1	2	2	1
Fig/Lit	1F	2F	2F	1F
Type	Holistic Multinary			
Category	Subordinate			

Other: **A person sitting down facing towards the right reading something in his hands.**

	NP (det N)	V prt	VP- V PP	VP- V N	PP
Phrase	1	1	2	2	1
Fig/Lit	1F	1	1L	2F	1F
Type	Holistic Multinary				
Category	Superordinate				

Literal:

A square at the top, a series of triangles in various positions in the middle,

	NP	PP	N	PP	PP- P adj N	PP
Phrase	1	1	1	1	1	1

and a rectangle at the bottom.

	(Conj) NP	PP
Phrase	1	1

Subject #107, stimulus set B, figure #1.

Self: **Situps**
N

Phrase 1
Fig/Lit 1F
Type Holistic Unitary
Category Distinguisher

Other: **Figure sitting down "knees" up and bent, square "head" curled in towards "knees"**

	N	V prt	N	adj adj	NP	V prt	PP
Phrase	1	1	1	2	1	1	1
Fig/Lit	1F	1F	1F	1L 1F	1L 1F	1F	1F

Type Holistic Multinary
Category Superordinate

Literal: **A shape with a block missing from the upper left side and bottom side,**

	NP	NP	adj	PP- P NP NP	
Phrase	1	1	1	1	1

vaguely a triangle.

	NP
Phrase	1

SAMPLE DATA: AUDIENCE COMPARISON

SELF	OTHER	LITERAL
A woman running	A person who is running towards the right	A square for the top. A corner of a triangle facing up & left and facing down & right. A long rectangle in vertical position and a small triangle to the left bottom and a small triangle to the right and bottom.

Phrases:	2	3	17
Fig/Lit:	2/0	2/1	
Nouns:	Subordinate Multinary	Superordinate Multinary	

(Set A #3, Sub 119)

REFERENCES

Akinnaso, F.N. (1982). On the differences between spoken and written language. Language and Speech, 25, 97-125.

Bartel, R. (1983). Metaphors and Symbols. Urbana, IL: National Council of Teachers of English.

Bilous, F.R. & Krauss, R.M. (1988). Dominance and accommodation in the conversational behaviors of same- and mixed-gender dyads. Language and Communication, 8, 183-194.

Bock, K. & Levelt, W. (1994). Language Production: grammatical encoding. In M.A. Gernsbacher (ed.), Handbook of Linguistics. pp 945-984. Academic Press: San Diego.

Bromley, D.B. (1991). Aspects of written language over adult life. Psychology & Aging, 6, 296-308.

Brown, R. & McNeill, D. (1966). The "tip of the tongue" phenomenon. Journal of Verbal Learning & Verbal Behavior, 5, 325-337.

Brownell, H.H., Potter, H.H., & Michelow, D. (1984). Sensitivity to lexical denotation and connotation in brain-damaged patients: A double dissociation? Brain and Language, 22, 253-265.

Brownell, H.H., Simpson, T.L., Bihrle, A.M., Potter, H.H., & Gardner, H. (1990). Appreciation of metaphoric alternative word meanings by left and right brain-damaged patients. Neuropsychologia, 28, 375-383.

Burbules, N.C., Schraw, G., & Trathen, W. (1989). Metaphor, idiom, and figuration. Metaphor and Symbolic Activity, 4, 93-110.

Burke, D.M. & Laver, G.D. (1990). Aging and word retrieval: selective age deficits in language. In E.A. Lovelace (Ed.), Aging and Cognition: Mental Processes, Self Awareness, and Interventions, pp 281-300. Elsevier: North Holland.

Burke, D.M., MacKay, D.G., Worthley, J.S., & Wade, E. (1991). On the tip of the tongue: what causes word finding failures in young and older adults? Journal of Memory and Language, 30, 542-579.

Cassirer, E. (1946). Language and myth. New York: Dover.

Chafe, W. & Tannen, D. (1987). The relation between spoken and written language. Annual Review of Anthropology, 16, 383-407.

Clark, H.H. (1992). Arenas of language use. Chicago: University of Chicago Press.

Clark, H.H. & Gerrig, R.J. (1983). Understanding words with new meanings. Journal of Verbal Learning and Verbal Behavior, 22, 591-608.

Clark, H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22, 1-39.

Collins, A.M. & Quillian, M.R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Elffers, J. (1976). Tangram. The ancient Chinese shapes game. New York: McGraw-Hill.

Fainsilber, L. & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. Metaphor and Symbolic Activity, 2, 239-250.

Ford, W.R., Weeks, G.D., & Chapanis, A. (1980). The effect of self-imposed brevity on the structure of dyadic communication. Journal of Psychology, 104, 87-103.

Fussell, S.R. & Krauss, R.M. (1989). The effects of intended audience on message production and comprehension: reference in a common ground framework. Journal of Experimental Social Psychology, 25, 203-219.

Fussell, S.R. & Krauss, R.M. (1989). Understanding friends and strangers: the effects of audience design on message comprehension. European Journal of Social Psychology, 19, 509-525.

Gerrig, R.J. & Gibbs, Jr., R.W. (1988). Beyond the lexicon: creativity in language production. Metaphor and Symbolic Activity, 3, 1-19.

Gibbs, Jr., R.W. (1994). Th e poetics of mind. Cambridge: Cambridge University Press.

Gibbs, Jr., R.W. (1994). Metaphors we live by. Cambridge: Cambridge University Press.

Gibbs, Jr., R.W. (1984). Literal meaning and psychological theory. Cognitive Science, 8, 275-304.

Gildea, P. & Glucksberg, S. (1983). On understanding metaphor: the role of context. Journal of Verbal Learning and Verbal Behavior, 22, 577-590.

Glantz, S.A. (1992). Primer of Biostatistics. New York: McGraw Hill.

Glucksberg, S. (1989). Metaphors in conversation: How are they understood? Why are they used? Metaphor and Symbolic Activity, 4, 125-143.

Glucksberg, S., Kreuz, R.J., & Rho, S.H. (1986). Context can constrain lexical access: implications for models of language comprehension. Journal of Experimental Psychology: Learning, Memory & Cognition, 12, 323-335.

Heilman, K.M. & Valenstein, E. (1985). Clinical Neuropsychology. New York: Oxford University Press.

Hoffman, R.R. (1984). Recent psycholinguistic research on figurative language. Annals of the New York Academy of Sciences, 433, 137-166.

Hunt, E. & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. Psychological Review, 98, 377-389.

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of codability and discriminability of the referents on the collaborative referring procedure. British Journal of Psychology, 82, 449-462.

Isaacs, E.A. & Clark, H.H. (1987). References in conversation between experts and novices. Journal of Experimental Psychology: General, 116, 26-37.

Kaplan, J.A., Brownell, H.H., Jacobs, J.R., & Gardner, H. (1990). The effects of right hemisphere damage on the pragmatic interpretation of conversational remarks. Brain and Language, 38, 315-333.

Katz, J.J. & Fodor, J.A. (1963). The structure of a semantic theory. Language, 39, 170-210.

Kemper, S., Kynette, D., Rash, S., O'Brien, K., & Sprott, R. (1989). Lifespan changes to adults' language: effects of memory and genre. Applied Psycholinguistics, 10, 49-66.

Krauss, R.M. & Glucksberg, S. (1977). Social and nonsocial speech. Scientific American, 236, 100-105.

Krauss, R. M. & Weinheimer, S. (1967). Effect of referent similarity and communication mode on verbal encoding. Journal of Verbal Learning and Verbal Behavior, 6, 359-363.

Lakoff, G. & Johnson, M. (1980). Metaphors we live by. Chicago: University of Chicago Press.

Lezak, M. (1983). Neuropsychological assessment. New York: Oxford University Press.

McCann, C.D. & Higgins, E.T. (1992). Personal and contextual factors in communication: a review of the 'communication game'. In G.R. Semin & K. Fiedler (Eds.), Language, Interaction, and Social Cognition, pp. 144-172. London: Sage Publications.

Mac Cormac, E.R. (1985). A cognitive theory of metaphor. New York: MIT Press.

Mac Whinney, (1989). Competition and lexical categorization. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), Linguistic Categorization, pp 195-242. Amsterdam: John Benjamins Publishing Co.

Martin, R.M. (1987). The meaning of language. Cambridge: MIT Press.

Mazzie, C.A. (1987). An experimental investigation of the determinants of implicitness in spoken and written discourse. Discourse Processes, 10, 31-42.

Mulac, A. & Lundell, T.L (1986). Linguistic contributors to the gender-linked language effect. Journal of Language and Social Psychology, 5, 81-101.

O'Brien, D.G., Higginson, B., & Miller, G.M. (1986). Factors affecting the interpretation of figurative language uses. Reading Research and Instruction, 25, 80-90.

O'Brien, D.G. & Martin, M.A. (1988). Does figurative language present a unique comprehension problem? Journal of Reading Behavior, 20, 63-87.

Oden, G.C., Ruecki, J.G., & Sanocki, T. (1991). Making sentences make sense, or words to that effect. In G.B. Simpson (Ed.), Understanding Words and Sentences pp285-303. Elsevier: North Holland.

Olson, D.R. (1988). Or what's a metaphor for? Metaphor and Symbolic Activity, 3, 215-222.

Pickens, J.D. & Pollio, H.R. (1979). Patterns of figurative language competence in adult speakers. Psychological Research, 40, 299-313.

Pillon, A., Degauquier, C. & Duquesne, F. (1992). Males' and females' conversational behavior in cross-sex dyads: from gender differences to gender similarities. Journal of Psycholinguistic Research, 21, 147-172.

Pinker, S. (1994). The language instinct. New York: William Morrow & Co.

Pollio, H.R., Smith, M.K., & Pollio, M.R. (1990). Figurative language and cognitive psychology. Language and Cognitive Processes, 5, 141-167.

Pratt, M.W. & Robins, S.L. (1991). That's the way it was: age differences in the structure and quality of adults' personal narratives. Discourse Processes, 14 73-85.

Purcell, W.M. (1990). Tropes, transsumptio, assumptio, and the redirection of studies in metaphor. Metaphor and Symbolic Activity, 5, 35-54.

Readence, J.E., Baldwin, R.S., & Rickelman, R.J. (1983). Word knowledge and metaphorical interpretation. Research in the Teaching of English, 17, 349-358.

Schober, M.F. & Clark, H.H. (1989). Understanding by addressees and overhearers. Cognitive Psychology, 21, 211-232.

Tannen, D. (1982). Oral and literate strategies in spoken and written narratives. Language, 58, 1-21.

Tannen, D. (1990). You just don't understand: women and men in conversation. New York: Ballantine.

Tartter, V.C. (1986). Language Processes. New York: Holt, Rhinehart & Winston.

Waggoner, J.E. (1990). Interaction theories of metaphor: psychological perspectives. Metaphor and Symbolic Activity, 5, 91-108.

Walker, V.G., Roberts, P.M., & Hedrick, D.L. (1988). Linguistic analyses of the discourse narratives of young and aged women. Folia Phoniatrica, 40, 58-64.

Wheeler, C.J. (1990). A question with no answer, or: reality as literalism and as metaphor. Metaphor and Symbolic Activity, 5, 55-62.

Whorf, B.L. (1956). Language, thought and reality. Cambridge, MA: MIT Press.