INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 8HR

THE STATUS OF CHOMSKY'S THEORY OF TRANSFORMATIONAL GENERATIVE GRAMMAR AS A MENTALIST THEORY OF LANGUAGE

by

JOHN S. RITTER

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

1978

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

6/8/78
date

D. Terence Langendoen
Chairman of Examining Committee

6/8/78
date

D. Terence Langendoen
Executive Officer

Professor Arthur Bronstein

Professor Arnold Koslow

Professor D. Terence Langendoen

Professor John Moyne

Supervisory Committee

Abstract

THE STATUS OF CHOMSKY'S THEORY
OF TRANSFORMATIONAL GENERATIVE GRAMMAR
AS A MENTALIST THEORY OF LANGUAGE

by

John S. Ritter

Adviser: Professor D. Terence Langendoen

This study examines the merits of Chomsky's claim that linguistics should be a mentalistic science and the merits of his claim that Transformational Generative Grammar is the called-for mentalist theory of language and language acquisition. Chapter 1 considers recent arguments that have been advanced in favor of an autonomous non-mentalistic linguistics and finds them inaccurate and unconvincing and concludes that the twin goals of providing grammars that are psychologically real and a general linguistic theory that can double as a model of the child's language acquisition device are both legitimate and desirable. Chomsky's claim, however, that the grammars that speakers internalize are transformational grammars is found incompatible with evidence from both psycholinguistics and automata theory. Chapters 2, 3, and 4 consider Chomsky's Language Acquisition Device. Chapter 2 concludes that the linguist's choice of a general linguistic theory is hopelessly

underdetermined by the available linguistic evidence. Chapter 3 concludes that the linguist lacks the evidence necessary to construct an evaluation measure and that notation-based evaluation measures make highly counter-intuitive claims about the set of possible human grammars. Chapter 4 concludes that even if we had an accurate enumeration of the set of possible human grammars and an evaluation measure, the Language Acquisition Device cannot be an explanatory model of language acquisition since it cannot, in principle, be speeded up to the point where it can operate in real time. Chapter 4 concludes that if we wish to develop an explanatory model of language acquisition we must seek an algorithm capable of directly constructing grammars for a language on the basis of the data. The claim that there is a single "correct" or "simplest" grammar which all speakers internalize and the language acquisition device must output is viewed not only as being false but as placing an unreasonable burden on a grammar construction algorithm.

ACKNOWLEDGMENTS

I would like to thank Professors D. Terence Langendoen, Arthur Bronstein, Arnold Koslow and John Moyne, the members of my dissertation committee, and Professor Samuel Levin, the outside reader, for giving unstintingly of their time and counsel. The members of the committee have at every phase read the dissertation promptly and carefully and at considerable cost to their already full schedules. The lengthy discussions of the contents of the dissertation were actually enjoyable (to me, at least) since the committee's suggestions for changes were appropriate, valuable, and offered in an entirely constructive spirit. It should not be assumed, however, that the members of the dissertation committee or the outside reader wholeheartedly endorse the entire slate of views espoused in this study. I would also like to thank Constantine Kaniklidis both for reading and commenting on the dissertation and for our numerous conversations over the years on formal linguistics. Even the most consummate blockhead could not have failed to benefit from his insights and encyclopedia knowledge. Finally, I would like to thank Madeline Ritter for being a highly critical first sounding board for many of these ideas.

TABLE OF CONTENTS

INTRODUCTION

This dissertation examines the merits of Chomsky's claim that linguistics should be a mentalistic science whose grammars are psychologically real and whose general theory can also serve as a model of the speaker's language acquisition device. It also seeks to determine whether Chomsky is correct in claiming that the grammars speakers internalize are transformational grammars or correct in claiming that Transformational Generative Grammar is a model of the child's language acquisition device.

The dissertation consists of four main chapters and a conclusion. Chapter 1 discusses whether or not linguistics should be a mentalistic science. Two major types of challenge to Chomsky's position will be considered: (1) recent attacks on Chomsky's claim that linguistics should be a mentalistic science by linguists who do not view linguistics as a subdiscipline of psychology and (2) recent attacks on Chomsky's claim that transformational generative grammar is a mentalistic science by linguists who feel that linguistics should be a subdiscipline of psychology but isn't. Chapters 2 through 4 examine the adequacy and testability of Chomsky's Language Acquisition Device (L.A.D.). Chapter 2 examines the testability of hypotheses concerning the form of Universal Grammar. Chapter 3 examines the adequacy and testability of claims made about the form and functioning of the evaluation measure. And Chapter 4 examines the operation of the Language Acquisition Device as a whole to

determine whether or not it is a realistic and explanatory model of language acquisition. Chapter 5 presents the conclusions of this study.

Since Chomsky's views have, in the past, been seriously misrepresented by some critics, the study makes every effort to present a comprehensive and accurate statement of his views--particularly those that will be called into question. The reader is thus in a position to evaluate fully any criticisms that are made and additionally in a position to uncover any inadequacies in Chomsky's position that may have been overlooked or ignored in this study.

CHAPTER 1

SHOULD LINGUISTICS BE A MENTALISTIC SCIENCE AND IS TRANSFORMATIONAL GENERATIVE GRAMMAR A MENTALISTIC SCIENCE OF LINGUISTICS?

In this chapter two major lines of attack on Chomsky's claim that linguistics should be a mentalistic science and his claim that Transformational Generative Grammar (T.G.G.) is the desired mentalistic theory of language are considered: (1) recent attacks on Chomsky's claim that linguistics should be a mentalistic science by scholars who do not feel that linguistics is best viewed as a subdiscipline of psychology and (2) recent attacks on Chomsky's claim that transformational generative grammar is a mentalistic science by scholars who feel that linguistics should be, but isn't, a subdiscipline of psychology. The chapter consists of three sections. Section 1 presents Chomsky's arguments (a) for the claim that linguistics should be a mentalistic science and (b) for the form such theory should take. Section 2 examines the attacks that have been made on the claim that linguistics should be a mentalistic science. Section 3 considers the attacks that have been made on the claim that Chomsky's T.G.G. is a mentalistic theory of language.

1.1 Chomsky's Conception of Linguistics as a Mentalistic Science

This section traces the evolution of Chomsky's theorizing about the goal of making linguistics a mentalistic science. It consists of three parts. Subsection 1.1.1 reconstructs the steps taken by Chomsky

in advancing the claim that the grammars linguists write should be psychologically real (i.e., isomorphic with a subpart of the grammatical system internalized and used by the speakers of a language) and describes the model of the speaker's hearer's linguistic abilities that was developed by Chomsky amd Miller. Subsection 1.1.2 presents Chomsky's claims concerning the implications of his research for psychology. Subsection 1.1.3 describes the evolution of Chomsky's thinking about the form and operation of the language acquisition device (a topic which will be discussed in greater detail in subsequent chapters).

1.1.1 On the Goal of Providing Grammars That Are Psychologically Real

The task of reconstructing the chain of reasoning behind Chomsky's claim that the goal of linguistics should be to provide grammers that are psychologically real and behind his claims concerning the actual structure of a model of the speaker-hearer's linguistic abilities is complicated by the fact that these goals underlay, but were neither described nor justified in two of Chomsky's earliest major works: The Logical Structure of Linguistic Theory (1955) and Syntactic Structures (1957). Chomsky explains his reluctance, in those works, to explicitly state these goals in his 1975 introduction to The Logical Structure of Linguistic Theory (LSLT): "In LSLT the 'psychological analogue' to the methodological problem of constructing linguistic theory is not discussed, but it lay in the immediate background of my own thinking. To raise the issue seemed to me, at the time, too audacious" (Chomsky 1955/75 p.35).[1]

The "psychological analogue" to the linguist's construction of a grammar that Chomsky had in mind was, of course, the child's acquisition of a grammar. The linguist has a corpus of data for which he is trying

to formulate a set of rules (i.e., a grammar) which the linguist must select on a principled basis from among a set of competing grammatical descriptions. The linguist's counterpart--the child--has a corpus of utterances which the child has heard and which must be explained by a grammar which the child selects from among a set of competing grammatical descriptions on some principled basis. If the linguist and the child select their grammar from the same pool using the same principle(s), both the task and the result will be strictly analogous (i.e., the child and the linguist will both arrive at the same grammar in the same manner). The general linguistic theory in this case becomes simultaneously an account of linguistic principles and methodology and an account of the principles and procedures the child has available innately to bring to bear in language acquisition.

Chomsky claims in his 1975 introduction to LSLT that it was an unstated assumption of LSLT that the linguist's results were the same as the child's:

> The 'realist interpretation' of linguistic theory is assumed throughout and it is argued that the competence attained by the normal speaker-hearer is represented by a transformational generative grammar, which determines the representation of each sentence in the levels of phrase structure and transformational structure (inter alia). These representations are then employed in the use and understanding of language, and provide the basis for the more general theory of language that will be concerned with meaning and reference, the conditions of appropriate use of language, how sentences are understood, performance in concrete social situations, and in general, the exercise of linguistic competence in thought and communication [Chomsky 1955/75 p. 45].[2]

Whatever the underlying assumptions of Chomsky's earlier writings may have been, his first published discussion of the goal of providing a linguistic model of the speaker-hearer's linguistic abilities appears in his 1958 paper "Linguistcs, Logic, Psychology and Computers." In this

work, Chomsky describes three abilities that the linguistic model must replicate: (1) the speaker's ability to distinguish sentences from non-sentences, (2) the speaker's ability (in most instances) to understand new sentences immediately and automatically, and (3) the speaker's ability to instantaneously produce desired sentences which may be entirely new to the speaker. To duplicate these abilities, we must construct the following three devices: (1) a grammar, (2) a device for determining uniquely and uniformly a structural description for each sentence (a speech recognition device),[3] and (3) a device that accounts for the speaker's ability to select a particular sentence (a speech production device). Chomsky maintains in this paper that in learning the language the child develops "something like a rather abstract theory to account for specific subject matter (observations of particular sentences)" (p. 433). The adult who recognizes and understands a sentence has "succeeded in determining that this sentence is 'predicted' (generated) by this theory, and in determining how it is generated" (p. 433). In speech production, the speaker "on the basis of inputs about which we know too little even to speculate . . . selects a set of 'choices' which, taken as inputs to the grammar, determine a particular generated sentence" (p. 431).[4]

It is, however, important to recall that while Chomsky views the grammar as playing a crucial role in speech production and recognition, he also warns the reader not to confuse the grammar with a production or recognition device. Even in Syntactic Structures (which did not so much as allude to the possibility of a mentalistic science of linguistics), Chomsky was careful to warn the reader against jumping to such a false conclusion:

> We have described these grammars as devices for generating sentences. This formulation has occasionally led to the idea that there is a certain asymmetry in grammatical theory in the sense that grammar is taking the point of view of the speaker rather than the hearer; that it is concerned with the process of producing utterances rather than the 'inverse' process of analyzing and reconstructing the structure of given utterances. Actually, grammars of the form that we have been discussing are quite neutral as between speaker and hearer, between synthesis and analysis of utterances. A grammar does not tell us how to synthesize a specific utterance; it does not tell us how to analyze a particular given utterance. In fact, these two tasks which the speaker must perform are essentially the same and are both outside the scope of grammars of the form (35). Each such grammar is simply a description of a certain set of utterances, namely, those which it generates [Chomsky 1975 p. 48].[5]

The description of the speech production and recognition devices becomes a bit more detailed in later papers but the descriptions always remain highly general and programmatic at best and Chomsky continually advises the reader that "it seems to me very likely that attempts to construct a model for the speaker or hearer are quite premature at this point, since we can hardly claim to have an adequate characterization of the form of the grammars that provide the devices that are employed, in some way, in the production and understanding of speech" (Chomsky 1961a p. 127).

The most detailed account of the sentence recognition device appears in Chomsky and Miller's 1963 paper "Finitary Models of Language Users":

> Instead of a relatively passive process of acoustic analysis followed by identification and symbolic representation, we imagined . . . an active device that recognizes its input by discovering what must be done in order to generate a signal (in some possibly derived form) to match it. At the heart of this active device, of course, is a component M that contains rules for generating a matching signal. Associated with M would be components to analyze and (temporarily) to store the input, components that reflect various semantic and situational constraints suggested by the context of the sentence, a heuristic component that could make a good first guess, a component to make the comparison of the input and the internally generated signals, and perhaps others. On the basis of an initial guess, the

> device generates an internal signal according to the rules stored in M and tests its guess against the input signal. If the match is unsatisfactory, the discrepancy is used to make a better guess. In this manner the device proceeds to modify its own internal signal until the match is judged satisfactory or the input is dismissed as unintelligible. The program for generating the matching signal can be taken as the symbolic representation of the input (Miller and Chomsky 1963 p. 465).

Chomsky and Miller consider the grammatical subcomponent (M) of the speech recognition device and the heuristic subcomponent (that generates the reasonable first guess) to be the most important. They feel, however, that we stand a good chance of making progress in our description of the grammatical subcomponent M. The grammatical subcomponent M is a device which "takes as its input a string x of symbols and attempts to understand it; that is to say, M tries to assign to x a certain structural description F(x) or a set $\{F_1(x), \ldots, F_m(x)\}$ of structural descriptions in the case of a sentence x that is structurally ambiguous in m different ways" (p.466). No provision was made in this paper for supplying a semantic interpretation of these structural descriptions since semantic theory was not sufficiently well developed at the time this paper was written (cf. p.466).

Chomsky and Miller propose that we divide the grammatical subcomponent M into two further subcomponents M_1 and M_2. M_1 utilizes short term memory to perform computations on the input string as it is received symbol by symbol. This "superficial" analysis is transmitted to M_2 which contains a transformational grammar stored in long-term memory. M_2 utilizes the full resources of the transformational grammar to determine the deep structure on the basis of the information provided by M_1.[6]

Chomsky and Miller suggest the following test of the psychological reality of their model:

> The psychological plausibility of a transformational model of the language user would be strengthened, of course, if it could be shown that our performance on tasks requiring an appreciation of the structure of transformed sentences is some function of the nature, number, and complexity of the grammatical transformations involved [Miller and Chomsky 1963 p.481].

A topic of grave concern to Chomsky and Miller is the question of whether or not human memory limitations should be incorporated into the form of their theory. The problem they are confronting is that

> . . . natural languages are not adequately characterized by one-sided linear grammars (finite automata), yet we know they must be spoken and heard by devices with bounded memory. . . . No automaton with bounded memory can produce all and only the grammatical sentences of natural language; every such device, man presumably included, will exhibit certain limitations [Miller and Chomsky 1963 p.421).

Chomsky and Miller are aware that many people will see a paradox in the fact that the human cognitive capacity is that of a finite state grammar, yet the grammars that Chomsky and Miller are postulating as part of the recognition device greatly exceed that in power (and may in fact correspond to the most powerful of the automata--the Turing machine). Chomsky and Miller feel, however, that the paradox is not real and there is no real need to incorporate human memory limitations into the form of human grammars. They claim (cf. p. 427) that while the device M may very well turn out to have insufficient computational space (i.e., memory) to permit it to understand all of the sentences predicted by the grammar to be grammatical, this turn of events has no greater significance for the form of our grammatical theory than the similar discovery that there are numerous mathematical calculations that humans cannot perform has for the form in which we write up our mathematical algorithms. In both cases the addition of memory aids (such as paper and a pencil) expands the class of calculations that humans can perform accurately. The fact that

the class of possible computations is expandable in this manner suggests that the memory limitations are not incorporated in the mathematical and grammatical algorithms themselves but rather are strictly a function of the amount of computational space humans have available for the use of these algorithms.

The existence of instances where humans fail to be able to recognize (or produce) sentences that are predicted by the grammar to be grammatical becomes problematical only when our predictions as to where the recognition (or production) device will "breakdown" (i.e., fail to recognize or produce a sentence) do not correspond to where linguistic performance actually breaks down. In other words, we have a problem when we predict on the basis of our knowledge of the computational space required by the grammar and the amount of memory we know humans to have available that a speaker will be unable to recognize a sentence (or produce it) and speakers do, in fact, recognize (or produce) the sentence in question, or when we predict that the speaker should be able to recognize (or produce) a sentence and speakers cannot. It is unfortunate thatpsycholinguists have contented themselves with checking to see if the familiar instances of human performance failures (e.g., with multiply embedded sentences) can be explained away in terms of the interaction of human memory limitations with the grammar. It may very well be the case that there are instances where we should expect a performance failure and no such failure occurs. If this were the case we would have to conclude that speakers are not performing their calculations with a device requiring expanded memory for its computations.[7]

Virtually nothing of substance has been added by Chomsky to

expand the description of the performance model since his 1963 paper with Miller. Chomsky did, however, clarify the distinction between competence and performance that was implicit in the 1963 paper. The classic description of the competence/performance distinction is, of course, Chomsky's Aspects discussion:

> Linguistic theory is concerned primarily with an ideal speaker-listener in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interests, and errors (random or characteristic) in applying his knowledge of the language in actual performance. . . . To study actual linguistic performance, we must consider the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one. . . . We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations) [Chomsky 1965 pp.3-4].

1.1.2 Implications for Psychology

Chomsky's first detailed discussion of the implications of his research for theoretical psychology appeared in his 1959 review of B. F. Skinner's Verbal Behavior. In this review, Chomsky maintains that language acquisition involves internalizing a highly complex and abstract formal grammatical system. This system is not learned primarily through instruction and careful differential reinforcement, but rather, much more casually and with remarkable rapidity. Chomsky reminds us of the common observation that the children of immigrants often learn their second language in the streets from other children with amazing rapidity. The speech of these children is, typically, completely fluent and correct to the last allophone; while their parents, no matter how high their motivation may be and how much instruction they may have received, seem completely to miss subtleties that have long since become "second nature" to their children. (cf. p.42).

Chomsky also rejected the assumption that the new sentences which we produce and understand are related to sentences we have previously used or encountered by any simple notion of formal, semantic, or statistical similarity or identity of grammatical frame. Chomsky charges that "talk of generalization in this case is entirely pointless and empty" (p.56). We produce and recognize new sentences because they are generated by the grammar we internalize and because we are capable of determining the process by which these sentences are derived in the grammar. Chomsky doesn't claim, however, that reinforcement and instruction play no role whatsoever in language acquisition. He maintains that:

> As far as acquisition of language is concerned, it seems clear that reinforcement, casual observation, and natural inquisitiveness (coupled with a strong tendency to imitate) are important factors, as is the remarkable capacity of the child to generalize, hypothesize, and 'process information' in a variety of very special and apparently highly complex ways which we cannot yet describe or begin to understand, and which may be largely innate, or may develop through some sort of learning or through maturation of the nervous system. The manner in which such factors operate and interact in language acquisition is completely unknown. It is clear that what is necessary in such a case is research, not dogmatic and perfectly arbitrary claims, based on analogies to that small part of the experimental literature in which one happens to be interested. The pointlessness of these claims becomes clear when we consider the well-known difficulties in determining to what extent inborn structure, maturation, and learning are responsible for the particular form of a skilled or complex performance . . . [Chomsky 1959 p.43].

1.1.3 The Language Acquisition Device

Chomsky's first reference in print to the goal of providing a language acquisition device appears in his 1958 paper, "Linguistics, Logic, Psychology and Computers." The details concerning the internal structure of this device were skimpy, but Chomsky was convinced that

> . . . the device . . . will have complex properties beyond the ability to match, generalize, abstract and categorize items in the simple ways that are usually considered to be available to the organisms.

> In other words, the particular direction that language learning follows may turn out to be determined by genetically determined maturation of complex "information-processing" abilities, to an extent that has not, in the past, been considered at all likely [Chomsky 1958 p.433].

The internal structure of the device is described in greater detail in "Explanatory Models in Linguistics" (1962a). Grammars are acquired by means of a language acquisition device which essentially incorporates the general theory of linguistic structure. In other words, it has available to it an advance specification of the form a grammar can take (i.e., an enumeration of the class of possible sentences, structural descriptions, and grammars, a method for determining the structural description of an arbitrary sentence for any of the permitted grammars, and an evaluation procedure). The goal of the language acquisition device is to select the most highly valued grammar that is compatible with the data.

Chomsky explains how the device operates in Aspects of the Theory of Syntax (1965). The language acquisition device has available to it a set of primary linguistic data.[8] The device then searches through a set of possible hypotheses G_1, G_2, ... (which are available through the algorithm that enumerates the set of possible human grammars) and selects those grammars that are compatible with the primary linguistic data (which is done by using the function that assigns a structural description to any arbitrary sentence to determine whether the grammar is compatible with the data). The device then selects the most highly valued grammar of this set using the evaluation measure.

This model is not, however, claimed to be totally realistic. Chomsky and Miller point out in their 1963 paper "Introduction to the Formal Analysis of Natural Languages" that

> Of course, we would have to supply the language-learning device with some sort of heuristic principles that would enable it, given its input data and a range of possible grammars, to make a rapid selection of a few promising alternatives, which could then be submitted to a process of evaluation, or that would enable it to evaluate certain characteristics of the grammar before others. The necessary heuristic procedures could be simplified, however, by providing in advance a narrower specification of the class of potential grammars. The proper division of labor between heuristic methods and specification of form remains to be decided, of course, but too much faith should not be put in the powers of induction, even when aided by intelligent heuristics, to discover the right grammar. After all, stupid people learn to talk, but even the brightest apes do not [Chomsky and Miller 1963 p.277].

Chomsky (1965) suggests that one possible heuristic strategy might be to consider only grammars that have better than a certain value. In addition he notes that

> What is required of a significant linguistic theory, then, is that given primary linguistic data D, the class of grammars compatible with D be sufficiently scattered, in terms of value, so that the intersection of the class of grammars compatible with D and the class of grammars which are highly valued be reasonably small. Only then can language learning actually take place [Chomsky 1965 p.203, fn.22].

One last idealization should be noted. Chomsky is describing "an idealization in which only the moment of acquisition of the correct grammar is considered" (Chomsky 1965 p.202, fn.19). A realistic model would have to account for the construction and abandonment of a succession of grammars of increasing sophistication, simplicity and accuracy by the child.

1.2 The Case for an Autonomous Linguistics

In the last five years, a small but growing number of scholars within the transformational school (as well as a larger number of scholars from without) have challenged Chomsky's mentalist conception of linguistics and called for a return to an autonomous ("unpsychologized")

linguistics.[9] Their arguments on behalf of this position will be presented and evaluated here. Unfortunately, several of the strongest and most carefully developed arguments against the mentalist conception of linguistics exist only in first draft form and cannot be attributed to their authors--who still have the option of revising or even abandoning them. To get around this difficulty I will present a composite picture of the anti-mentalist arguments that could potentially be made without attributing them to a specific author or authors. None of the arguments will be of my own devising, so in no instance will we be evaluating a straw man. In fairness to the scholars who have advanced the autonomous linguistics position, it should be noted that no single scholar has made or endorsed the entire slate of anti-mentalist arguments presented here. It is, in fact, likely that each would find at least one argument that he wouldn't care to endorse.

The autonomists maintain that the proper subject matter for linguistic investigation is language--not knowledge of language. They ask why linguists should concern themselves with how humans acquire knowledge of their language or with the specific form in which such knowledge is represented when such disciplines as mathematics, physics, and economics are perfectly content to leave research into these topics in the hands of psychologists. They question whether Chomsky's enlarged domain of linguistic inquiry is any more efficient, productive or "natural" than the traditional division of labor between linguistics and psychology. Several of the autonomists suggest that if linguistics is to be incorporated into another discipline, it should be mathematics, not psychology.

The analogy with mathematics has inspired many of the anti-mentalist arguments that have been advanced. Several of the autonomists have claimed that arguments against an autonomous linguistics should be translatable, by parity of reasoning, into arguments against an autonomous mathematics. They conjecture that no mathematician could be convinced that the proper goal of that discipline is to replicate the ideal human mathematician's knowledge of a number system and explain how humans acquire such knowledge.

One of the autonomists maintains that the mentalist/autonomist controversy in linguistics can be clarified and possibly resolved if we re-examine the Nominalist/Conceptualist/Platonist debate that took place in mathematics. The Nominalist view of mathematics (propounded by J. S. Mill) is that mathematical propositions represent truths about how collections of discrete physical objects can be separated and rearranged. We know, for instance, that 1 + 1 = 2 because our manipulation of physical objects reveals this to be the case. Arithmetic and geometric propositions are, thus, inductive generalizations about facts observed in nature (i.e., they are empirical propositions like those we are familiar with in physics, chemistry, etc., except that they are so confirmed by experience that we view these propositions as having certainty).

The Conceptualist view (espoused by Brentano, Erdmann and others) is that arithmetical and geometrical propositions represent psychological laws about human mathematical reasoning. The equation 1 + 1 = 2 is true because that is the solution supplied by the operation of our mental mathematical faculty. The goal of mathematics, according to this viewpoint, is to produce an axiomatization of human mathematical reasoning.

The Platonist view of arithmetic (endorsed by Frege and many other mathematicians) is that the subject matter of mathematics is not arrangements of physical objects or psychological laws of mathematical reasoning but rather numbers with are abstract objects.

The Nominalist view is described as having the following three problems: (1) according to this view it is, in principle, possible for empirical counterexamples to refute standard mathematical truths, yet the fact is that no one would accept such a refutation, (2) we view all mathematical truths as having the same degree of certainty even though we have not actually encountered any empirical support in the real world for most of these mathematical truths, and (3) it is impossible to supply a physical interpretation for numbers like $\sqrt{-1}$.

The Conceptualist view also allows for the possibility that standard mathematical truths may be empirically disconfirmed, say, if psychological testing reveals that humans add 88 and 8 and get 97 as a solution. Frege, a particularly insightful critic of both the Nominalist and Conceptualist views of mathematics, is quoted on the defects of the Conceptualist view:

> . . . it is impossible to ascribe to every person his own number one; for in that case we should first have to investigate the extent to which the properties of these ones agreed, and if one person said 'one times one is one' and the next said 'one times one is two', we would only register the difference and say your one has one property, mine has another. There could be no question of any argument as to who was right, or of any attempt to correct anyone; for they would not be speaking of the same object. Obviously, this is totally contrary to the sense of the word 'one' and the sense of the sentence 'one times one is one'. Since the number one, being the same for everyone, stands apart from everyone in the same way, it can no more be researched by making psychological observations than can the moon. Whatever ideas there may be of the number one in individual souls, they are still to be carefully distinguished from the number one, as ideas of the moon are to be from the moon itself [Frege 1967 p.16].

The linguistic analogues to these views are as follows: occupying the Nominalist position in linguistics is "taxonomic" structuralist linguistics which concerned itself with making inductive generalizations about the segmentation and distribution of physically objective linguistic units. Linguistics has two variants of the Conceptualist position which are called "Competencism" and "Performancism." Both views concern themselves with the speaker-hearer's knowledge of the language. The competencists (e.g., Chomsky) demand that the grammar be psychologically real only in the sense that it represents all and only the knowledge speaker-hearers have of their language and that the general theory be psychologically real only in the sense that it represents all and only the knowledge that children have initially about the possible form of grammars and the procedure for selecting one. The Performancists (e.g., Derwing, Watt, and in their recent writings Wayne Garrett, Janet Fodor and Jerry Fodor) make the stronger demand that grammars be considered psychologically real only if the rules and distinctions of the grammar play a role in the explanation of the on-line linguistic computations connected with speech production and perception and that general linguistic theory be considered psychologically real only if the distinctions, etc., it incorporates play a role in the explanation of the on-line computations taking place in language acquisition.

The Platonist (or autonomist) view of linguistics is that linguistics is not concerned with the knowledge speakers have of either their language or the general form of language, but rather, with the language and the linguistic universals themselves. Linguistic theories are, thus, concerned with abstract objects whose reality exists independent

of individual speakers. The goal of the linguist is to produce the optimal formal system which predicts and explains every property and relation of every sentence. These theories reproduce speaker intuitions but are not vulnerable to psychological or neurological disconfirmation.

The author of this very extended analogy accepts Chomsky's attack on the Nominalist position in linguistics and concentrates on arguing against the Conceptualist position. Two major arguments are invoked against (both variants of) the Conceptualist position. The first argument alleges that the Conceptualist position makes counter-intuitive claims about what constitutes being a speaker of a natural language. Consider a situation in which we are visited by aliens who are able to communicate fluently with us in English. We would naturally want to claim that these aliens are speakers of English. But suppose, as would be likely, a psychological or neurological examination of the aliens' brains revealed that they have a vastly different brain structure from our own and that the language processing mechanisms that they are using to communicate with us are completely different from our own except they agree with respect to the structural descriptions they assign to sentences. The grammar that is psychologically real for them is not the grammar that is psychologically real for us. Given such a situation, it is claimed in this argument that the conceptualist would be forced to claim, contrary to our intuitions, that the aliens are not speakers of English. Moreover, it is claimed that if the human speakers of English underwent some dramatic overnight change in their neurological structure while continuing to assign the same structural descriptions to sentences, the conceptualist would, nonetheless, be forced to claim that these people no longer speak English.

The second argument claims that grammatical rules are no more subject to psychological or neurological disconfirmation than are mathematical truths since they are both properly construed as being concerned with abstract objects. The linguist needn't abandon any of the rules of his/her optimal grammar just because psychological testing reveals that speakers do not have such a rule. The goal of the linguist should be to capture and explain the facts about language, not to duplicate speakers' possibly non-optimal accounts of these facts.

The Platonist approach is said to free the linguist from the Nominalist and Conceptualist constraints on the form of grammars that overly restrict their abstractness and prevent the linguist from writing optimal explanatory grammars. The removal of these restrictions, and the adoption of the Platonist approach in general, are said to cost us none of the advantages of the Conceptualist approach (e.g., its use of generative grammars, its use of multiple levels, and its ability to express abstract linguistic universals, etc.).

Another of the autonomists has drawn on mathematics for the following analogy: It is claimed that humans perform their mathematical calculations using an algorithm but lack an axiomatization of the mathematical principles that underlie and explain the effectiveness of these algorithms. The mathematician's concern is said to be with axiomatizations and the psychologist's with the algorithms. By analogy, linguists are concerned with axiomatizations and psychologists with the algorithms speakers use to encode and decode their language.

Chomsky has not responded to any of the anti-mentalist arguments just presented. This is not surprising since they have all arisen within

the last five years and, as indicated earlier, several of the papers in which these arguments are presented are presently in first draft form only and have not been widely circulated. Chomsky has, in fact, presented little in the way of justificatory arguments for his claims that linguistics should be a mentalistic science. Chomsky's major justification for adopting the mentalist approach is that if we focus attention on grammars that are psychologically real and attempt to specify a set of universals which define the terms "possible human grammar" and "possible human language" we will be in a position to draw highly significant conclusions about the nature and functioning of human intelligence and about how humans learn. At the same time, we would be significantly advancing our understanding of human language. Chomsky adopts, at certain points, a laissez-faire attitude with respect to studies that focus on other questions about language or its use and even with respect to studies that wish to make no claims about psychological reality. Thus, in his 1975 introduction to The Logical Structure of Linguistic Theory, Chomsky indicates that

> . . . there is no question of 'right or wrong'. It is merely a question of where one's interests lie. If someone prefers not to adopt a 'realist interpretation' of linguistic theory, and thus to make no claim for the empirical validity of the theoretical principles he adopts, I see no argument that could demonstrate to him that this conception must be abandoned. Or conversely [Chomsky 1955/75 p.39].

Chomsky's laissez faire attitude is obviously not shared by the autonomists--many of whom maintain that the shift to a mentalistic conception of linguistics entails a conceptual confusion of knowledge and the object of that knowledge and entails a loss of explanatory power. I believe, however, that Chomsky's view of the issue is substantially

correct and that the anti-mentalist arguments miss their mark. I will begin my evaluation of these arguments with a consideration of the autonomist and mentalist divisions of linguistic labor.

The autonomist division of labor sees the work of the psychologist, linguist and mathematician as being proper subsets of each other. The psychologist is concerned with the set of possible human grammars for the set of possible human languages, the autonomist linguist is concerned with the set of all possible grammars (including grammars that could not be internalized or used by humans) for the set of all possible human languages. The mathematician is presumably free to range over the entire set of grammars (human and nonhuman) for the entire set of languages (natural and artificial). The mathematician would no more accept being confined to considering artificial languages and their grammars than the autonomist linguist would accept being confined to working on psychologically unreal (i.e., nonhuman) grammars for human languages. The mentalist division of labor envisions the linguist and the psychologist as jointly investigating the set of possible human languages and the set of possible human grammars for them. The mathematician remains free to range.

We are now in a position to ask what there is about the mentalist's division of linguistic labor that forces the mentalist to claim that aliens who converse with us in English are not speakers of English while the autonomist linguist is free with his division of linguistic labor to claim that these aliens are, in fact, speakers of English. Actually there doesn't seem to be anything in either division of labor that would force the linguist to claim that the aliens are not speakers

of English. The mentalist is perfectly free to claim that anyone who possesses an adequate grammar of English (human or nonhuman) is a speaker of English. The mentalist both can, and would want to, define the English language as the set of sentences (and their structural descriptions) English is standardly considered to be comprised of and the mentalist both can, and would want to, define a speaker of English as any entity that has an internalized grammatical device which can produce and recognize all and only the set of English sentences (and their structural descriptions). No harm emanates from there being nonhuman speakers of a natural language. Thementalist is not concerned as a professional with them nor with their grammars of natural languages. As to the speakers of English who might undergo a sudden massive neurological change, the proper mentalist response is not that these individuals are no longer speakers of English but rather that the class of possible human grammars has changed.

Some autonomists might also be expected to be disturbed by the fact that the mentalist is countenancing two different grammatical descriptions for the same language. If there is a single optimal ("true") grammar for any language, then, it would mean that at least one of the species has a nonoptimal ("false") theory of the English language. The autonomist who wants to interpret this fact as having adverse consequences for the mentalist approach must first demonstrate that there is at most a single optimal ("true") grammatical description for every natural language. As we will see in Chapter 3, this is a highly implausible claim. One would also like to see the term "optimal" defined in terms of a set of nonconflicting criteria that the linguistic community can

agree on. Thus far, the adjective "optimal" has been bandied about without any attempt at definition whatsoever.

The claim that it is conceptually confusing for the linguist to study knowledge of language rather than language is, itself, confused. The linguist is studying both language and the speaker-hearer's knowledge of that language. Surely it is possible for the same person to study an object and knowledge of that object without any inherent confusion. The search for a psychologically real grammar is a search for knowledge of the object language as well as a search for knowledge about the speakers' knowledge of their language as long as the speakers' grammar represents an accurate account of the language they speak. The claim that speakers have accurate knowledge of the structure of their language does not seem particularly bold--despite the attempt of at least one autonomist to convince us that speakers have knowledge of how to manipulate their language but not knowledge about the language itself.

The claim that there is a wide chasm between the concerns of linguists and those of psychologists, because linguists are (or should be) concerned with axiomatizations (which reveal underlying principles) while psychologists are concerned with algorithms (which purportedly get the job done but without revealing the underlying principles that make the algorithm work), is highly problematical. Theproblem is that the distinction being promulgated between axiomatizations and algorithms is artificial. While it is true that specific algorithms may not be revealing, there is nothing that in principle prevents algorithms from revealing underlying principles. In fact, that is the goal of virtually everyone who writes algorithms for anything but the most trivial purpose. On

the other hand, there is no guarantee that an axiomatization will reveal the relevant underlying principles. Actually, there is nothing that can be captured (or revealed) by an axiomatization that is in principle uncapturable in the form of an algorithm and vice versa.[11] Any distinction drawn between the subject matters of linguistics and psychology on the basis of a claim that speakers are saddled with unrevealing algorithms is false. In addition, no evidence has been provided to support the claim that speakers construct unrevealing accounts of their language whether in the form of algorithms or axiomatizations--providing such evidence is clearly the first step to be taken in advancing a distinction of this sort.

The argument that linguistic rules are no more disconfirmable by psychological or neurological testing than mathematical propositions are, also fails. The conceptualists in mathematics claim that mathematical propositions are about human mathematical reasoning. The conceptualists in linguistics claim that linguistic rules are about language. The conceptualist linguist who learns that some rule is not psychologically real or for that matter learns that an entire grammar is not psychologically real is not forced to claim that the rule or the grammar is an inaccurate or false account of the language the way that the conceptualist mathematician would be forced to conclude that a psychologically disconfirmed mathematical proposition is false. All the conceptualist linguist is forced to conclude is that the rule or grammar in question is not the form in which speakers represent their knowledge of their language. Psycholinguistic tests are viewed by the conceptualist linguist as revealing something about speakers' knowledge of language, not

something about language. There is no fallacy, and nothing counter-intuitive, in claiming that psycholinguistic testing can disconfirm claims about the specific form of human Knowledge.

Still another dubious claim is the claim that "Platonic" linguistic theories remove limitations on abstractness that prevent linguists from producing the optimal (or "true") grammar of the language, a grammar whose optimality the grammars of humans can only approximate. There are many linguists who wonder if there are grammars outside the class of possible human grammars that are, in any significant sense, more optimal than those within. Certainly, this has yet to be demonstrated. Just how important are the differences between grammars that assign the same structural descriptions if we are not adopting a "realist interpretation" of the status of these grammars? If we have, for example, a large set of clocks--all agreeing on the time but differing in their internal structure--can we seriously claim that one clock (or even a small subset of the clocks) is providing the most revealing and true account of the time while the others are not? We can admittedly find other grounds for preferring one clock over another (e.g., for aesthetic reasons, because the clock has a long-lasting battery, because the clock is electric, etc.), but will there be a clock (or even a small set of clocks) that is optimal under every conceivable criterion or will there be conflicting criteria?[12]

The linguistic evidence suggests that we do not require any further loosening of our restrictions on linguistic descriptions to account for the facts of language. We would require the full power of unrestricted rewriting systems if human languages are strictly recursively

enumerable sets.[13] If a language is strictly recursively enumerable, we will be able to identify a string as being grammatical in a finite number of steps but there will be some strings (in fact, an infinite number of them) which we will be unable to identify as ungrammatical in a finite number of steps. But no evidence has ever been advanced to the effect that there are any ungrammatical strings which speakers are in principle incapable of identifying as such. In fact, while there have been quite a few claims that there are properties of natural languages that cannot be handled by context-free grammars, no one has ever claimed to have found some property of natural language which cannot be handled by context-sensitive grammars(which can at most generate recursive sets). While it is true that transformational grammars (of the Aspects type and all subsequent versions) are equivalent in power to unrestricted rewriting systems (or Turing machines), none of the grammars linguists have written has made crucial use of this power. Stan Peters has pointed out in several places that the grammars linguists have written have exponentially bounded cycling functions and are therefore generating recursive sets. Until autonomists find some property of natural language that requires the power of an unrestricted rewriting system, we have no reason to use such systems and every reason not to, since there are grammars in that class that cannot, even in principle, be learned by an algorithm.[14] If a grammar cannot in principle be learned by any algorithm, then that grammar is not a possible human grammar. If all the grammars for a language are unlearnable (and there are many instances where this is the case), then that language is unlearnable and cannot be considered a possible human language. Even the autonomists are claiming to be restricting their attention to the set of possible human languages.

Finally, it is not clear that the move to an autonomous ("Platonist") linguistics does not cost us some of the advantages of the mentalist ("Conceptualist") approach. For instance, the conceptualist approach defines the set of possible human languages in terms of the set of possible human grammars (the set of possible human languages is the set of languages generated by the set of possible human grammars--which is itself generated by algorithm). The autonomist is presumably not concerned with defining the set of possible human grammars. It remains to be seen, then, how the autonomist is going to be able to decide on the precise contents of the set of possible human languages without determining the set of possible human grammars (i.e., without reintroducing the goal that they want to ship back to psychology). The autonomous linguist is sure to find numerous instances where there is disagreement about whether or not some set of sentences is a possible human language. In a case like this, the conceptualist can hope to resolve the conflict by bringing to bear evidence from psycholinguistics about the possible form of a human grammar. If the language in question cannot be generated by a grammar of the specified form, then it is not a natural language. But where does the autonomous linguist go for additional evidence that can resolve the dispute?

None of the anti-mentalist arguments that has been considered here establishes that the idea of a "mentalist linguistics" is conceptually confused. Nor do there seem to be any strong arguments for seeking more powerful, less restrictive grammars. There is no reason, then, why linguists should not attempt to develop "psychologically real" grammars or a "psychologically real" general linguistic theory which explains

how speakers acquire these "psychologically real" grammars.

There are, in fact, some linguists who claim that the only worthwhile goal for linguists is the goal of providing "psychologically real" grammars and a "psychologically real" general linguistic theory. Derwing (1973) is particularly vehement on the subject:

> Unless some rational basis can be established for making us think that some very close relation holds between the linguist's (ultimate) grammar of any language L and the native speaker's own internalized 'mental' grammar for this same L, all the linguist's grammar-writing activity must surely degenerate into a kind of highly intellectualized and complex game, the results of which can seemingly have no relevance to anything in the real world and hence can be of no interest to anyone (except, perhaps, to that special type of individual who is simply drawn by natural inclination to crosswords, cryptograms, chess and other such intellectual puzzles for the pleasure of the mental stimulation which they afford) [Derwing 1973 p.47].

Derwing's disdain for research on nonhuman grammars for human languages seems excessive. It seems reasonable, though, to assume that the majority of linguists will prefer to direct their research toward making a contribution to our understanding of the nature of the human mind and its linguistic capabilities if this goal is at all achievable. Surely "a," if not "the," primary reason for studying human languages is to learn about the humans who are the source of our object of inquiry. But, as Chomsky admits, it is a question of interests and research priorities, not one of "right or wrong."

1.3 On Chomsky's Claim that Transformational Generative Grammar is a Mentalistic Theory of Language

The second line of attack directed against Chomsky's Transformational Generative Grammar (T.G.G.) takes umbrage not with the claim that linguistics should be a mentalistic science but rather with the claim that T.G.G. represents a mentalistic science of language. To see why

the charge has been leveled that T.G.G., despite its mentalist rhetoric and ideology, is not a mentalistic theory, it is necessary to consider the performance model developed by Chomsky and Miller. The production device, as described in Chomsky's 1958 paper, "Linguistics, Logic, Psychology and Computers," consists of two components: (1) a component which takes a set of inputs ("about which we know too little even to speculate" [p.431]) and selects a set of "choices" (i.e., a particular sentence to be produced) and (2) a transformational grammar which operates on the output of the first component producing as its output its particular desired sentence. The recognition device recognizes a sentence by "determining that this sentence is 'predicted' (generated) by this theory [i.e., the transformational grammar] and in determining how it is generated" (p.433). Chomsky and Miller (1963) describe the recognition devices as having two major components: (1) a set of heuristics which, using short-term memory, produce a reasonable first guess at the structure of the input string and (2) a component containing the transformational grammar in long-term memory which recognizes the input string by discovering whether and how it can be generated by the grammar. Chomsky and Miller suggest that the plausibility of this model could be tested by checking to determine if "our performance on tasks requiring an appreciation of the structure of transformed sentences is some function of the nature, number, and complexity of the grammatical transformations involved" (p.481). Despite the fact that several of the components of the performance model which perform critical operations were described in only the vaguest terms and in advance of the psychological tests they advocated, Chomsky and Miller claim that any satisfactory

model of performance will incorporate a transformational grammar as a vital component.[15]

Psycholinguists seeking to determine whether or not transformational grammars play any direct role in speech recognition have considered two ways in which the rules of a transformational grammar could be utilized. The most commonly considered method is called analysis by synthesis. It is the method most similar to the one described by Chomsky and Miller. Analysis by synthesis involves searching through the output of the transformational grammar to determine whether or not the grammar generates a particular input string. To keep the search from being impossibly lengthy a powerful set of heuristics is needed to cut down drastically the size of the search. They are said to accomplish this by providing a pre-analysis of the input string which defines a set of properties to be searched for and excludes a set of properties that need not be searched for. The second model which makes direct use of the transformational grammar is called analysis by analysis. It calls for providing a derivation by having bi-directional rules (something not easily definable for transformational rules) that permit us to run the grammar backwards, producing all the intermediate derivational stages until we return to the start symbol S. This approach also requires some sort of pre-analysis of the input string (i.e., an initial guess as to the surface structure of the string) to have the vaguest semblance of plausibility. No serious attempt has been made to provide a working model of either of these two types. Attempts have been made, however, to test both models. The reason that is considered feasible despite the lack of the heuristic "pre-analysis" components both models require, is that

both models predict that our performance on tasks requiring an appreciation of the structure of transformed sentences will be some function of the "nature, number, and complexity" of the grammatical transformations involved. This prediction is called the Derivational Theory of Complexity.

There have been numerous tests of the Derivational Theory of Complexity. The results of these tests are surveyed in Fodor and Garrett (1966), Bever (1968), Watt (1970), and Fodor, Bever, and Garrett (1974). The results of the experimental tests fail to support the claim that transformational derivations play any role whatsoever in actual sentence recognition. These results are accepted by virtually all psycholinguists--including those who had a vested interest in confirming that transformations are employed directly in the process of sentence recognition. There has been very little work done on the sentence production version of the Derivational Theory of Complexity but no one seems to be holding out much hope that the results will be any different in that quarter.

There is, of course, one form of challenge to which these experiments are all vulnerable--they are all based on specific theories of transformational grammar and within those theories based on specific formulations of rules. Anyone who has followed Chomsky's rapid passage through a succession of significantly different variants of T.G.G. realizes what a fast-moving target he represents for those who are intent on empirically testing the claims he has made. What we need, then, are results that are generalizable to all forms of transformational grammar. Fortunately, there are results from automata theory that may allow us to

make generalized claims about the plausibility of the analysis by synthesis and analysis by analysis models.

The results that I refer to have a bearing on the "heuristic" components that are needed to speed up the searches that are performed in the analysis by synthesis and analysis by analysis models of sentence recognition and are needed to speed up the techniques for "homing in" on the specific sentence we want our random sentence generating transformational grammars to produce. The search process for both recognition and production will be hopelessly inefficient and time consuming unless a set of "heuristics" exists which can speed up the process remarkably.

The first problem to be confronted in providing a set of heuristics to speed up the search activities of the recognition and production devices is a terminological one. Luchins and Luchins (1965) point out that "a heuristic procedure may be contrasted with an algorithm in the sense that the latter, if used correctly, invariably leads to solution, whereas a heuristic is not guaranteed to lead to solution but may help; it is used for its suggestive value" (p.136). Clearly what we need in connection with speech production and recognition is an algorithm, not a set of heuristics. Chomsky's claim is that speakers can, in principle, produce and recognize every sentence of their language. In other words, if temporal and memory limitations and other factors such as disinterest, etc., were removed, there would be no sentence that speakers could fail to be able to produce or recognize. If, however, we were talking about a situation in which heuristics were employed in sentence production and recognition, there would be some sentences of the language which could never be produced or recognized. This claim has never been made in

connection with these devices. We are searching, therefore, for an algorithm, not a set of heuristics. It is important to clear up this confusion because many psycholinguists have been misled into thinking they are searching for a loosely interrelated patchwork of heuristics that allows the linguist to get a job done in a quick and inelegant manner that could be accomplished in a more tedious and more revealing manner without such heuristics.

Once we have agreed that we are searching for an algorithm to speed up the arrival at a derivation for an input string or a desired output string, we are in a position to apply some results from automata theory. E. M. Gold (1967) has shown[16] that there is no algorithm that can consistently arrive at a target grammar, derivation, etc., faster than the enumeration algorithm. In other words, while another algorithm may arrive at some specific target grammars, derivations, etc., faster then the enumeration algorithm, it is not universally faster, so there must be some (possibly an astronomical number of) target grammars, derivations, etc., that are reached only after the (astronomical) amounts of time typically consumed by the enumeration algorithm. Miller (1965), for instance, estimated that the enumeration needed to recognize any given twenty-word string would require as much time to enumerate as the universe is currently estimated to have been in existence. It follows, then, that Chomsky's production and perception models cannot have a "heuristic" component that will make use of a transformational grammar in an efficient and realistic manner. Whatever the production and recognition devices are that are employed by humans, they cannot make direct use of a transformational grammar in their operation.

How, then, do we interpret the results of the psychological tests and the findings in automata theory? Fodor and Garrett (1966) proposed that "an acceptable theory of the relation between competence and performance models will have to represent that relation as abstract, the degree of abstractness being proportional to the failure of formal features of derivations to correspond to performance variables" (p.152). Fodor, Bever, and Garrett (1974, pp.369-71) present five possible responses to the fact that the grammar is not concretely realized in the perceptual model (i.e., the decoding processes compute the same structural descriptions as the grammar but the information they use in their computations is not represented in the same form as that in which it is represented in the grammar).

The first possible response to the question of what possible status the grammar and linguistic universals can have if they are not part of the production-recognition system is, of course, to deny that the evidence is strong enough to support such a conclusion. The evidence is, as we noted, not totally unassailable. Most psycholinguists (including Fodor, Bever, and Garrett) feel that the evidence is not about to be set aside. The independent evidence from automata theory which has yet to be considered by psycholinguists makes the conclusion quite convincing.

The second response Fodor, Bever and Garrett discuss concerns the status of linguistic universals. There are several possible interpretations for the facts here. We can claim that linguists were wrong about their proposed universals, or claim that they should be reinterpreted as properties of structural descriptions, or claim that the universals describe the production or recognition systems. The empirical status of linguistic universals will be considered in the next chapter.

The third possibility the authors discuss is that the universals don't apply to the recognition device. This device contains heuristics which do not always assign correct structural descriptions, and when this happens a "brute-force" analysis by synthesis routine actually utilizing the transformational grammar is used. The grammar is thus psychologically real even though it doesn't typically play a role in the language-processing system. It is difficult to envision how this is to work. Apparently the heuristics periodically fail (as by definition they should) and the "brute force" response to the situation is to employ a technique that is hopelessly impractical. But if this were the case, conversation would be punctuated with pauses the length of the existence of the universe and this is just not the case.

The fourth possibility is that

> Everyone is right. In this view, both the grammar and the processing heuristics are psychologically real. The function of the grammar is to provide a 'library' of information about the structure in a language, and the functioning of (some of the) heuristics is to access the grammar, i.e., to ask what the grammar says about the structure of the particular string of morphemes to which the hearer is attending. Such a relation between the grammar and the perception system is compatible with the negative evidence on DTC so long as the heuristics access the grammar by some means other than using it to actually run through grammatical derivations. If, in fact, there is an objection to this approach, it is only that there is, at present, no positive reason for believing it is true, and such sentence-processing heuristics as are supported by the available data are not of this kind [p.31].

Fodor, Bever and Garrett are, however, overlooking the fact that there is no method of access to the grammar faster than the enumeration technique. There is just no point in positing a library whose information cannot be had in a humanly relevant timespan.

The final possibility they suggest is that recognition procedures may only be constructable by an algorithm operating on grammars

that correspond to the standard linguistic universals in format. The child first learns a grammar of this sort and then constructs the recognition device. There is no evidence that this is the case. The burden of proof rests on the holder of such a position to demonstrate first that there are recognition procedures that cannot be devised directly but only via a transformational grammar of whatever sort Chomsky is presently advocating and second that humans use such a recognition procedure (since there will surely be procedures that produce recognition algorithms directly).

The position I prefer is that of Steinberg (1970):

> If a person is presumed to have one order of competence rules for production and another for understanding, why would he still need another order, that of Chomsky's model of competence. The postulation of the existence of an organization of language knowledge such as Chomsky's is theoretically superfluous [Steinberg 1970 p.187].

When evaluating the need for a "competence" grammar, we should keep in mind the fact that the production and recognition devices are highly abstract algorithms (not "heuristics") on the same order of "complexity," "systematicity," "formalness," etc., as Chomsky's "competence" transformational grammars (which are also algorithms). We should also keep in mind that if we do decide to concern ourselves exclusively with production and recognition devices, we are not committing the fallacy of dealing with linguistic performance rather than competence. Recall Chomsky's Aspects definition of the competence/performance distinction:

> Linguistic theory is concerned primarily with an ideal speaker-listener in a completely homogeneous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distraction, shifts of attention and interest, and errors (random or characteristic) in

> applying his knowledge of the language in actual performance. . . . We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations [Chomsky 1965 pp.3-4].

The production and recognition devices are just as much an idealization as the so-called "competence" grammars. These devices are not concerned with individual speakers and they are not concerned with actual linguistic performance. They are concerned with the knowledge speakers have of their language which is represented in the form of a production and a recognition device. The linguist who is interested in constructing the production and recognition devices is not going to be any more interested in such "irrelevant conditions as memory limitations, distractions, shifts of attention, and errors (random or characteristics) in applying his knowledge of the language in actual performance" than Chomsky is when writing a transformational grammar. Performance errors will be explained away in terms of factors outside of the purview of the production and recognition devices. The term "competence" should never have been misleadingly restricted to applying to the component of the "performance" model consisting of the transformational grammar. The entire "performance" model can be called the "competence" model since no component of the model is any more concerned with actual performance or any less idealized than the transformational grammar.

One can conclude, then, that the burden of proof is on the linguist who wishes to maintain that the recognition and production devices are derived by algorithm from a transformational grammar or who wishes to maintain that the production and recognition algorithms so abysmally fail to explain the facts of language that we must assume that speakers construct an additional algorithm (the transformational grammar) to

rectify this inadequacy. Until we see concrete evidence that either of these situations obtains, we have no reason to maintain that transformational grammars are in any sense "psychologically real," and it would be dogmatism (of the type Chomsky attributes to Skinner) to claim otherwise.

I have yet to present Chomsky's response to the negative results of the psycholinguistic tests. Chomsky has not endorsed Fodor, Bever and Garrett's efforts to salvage the claim that transformational grammars are in some sense "psychologically real." In his 1976 paper, "On the Biological Basis of Language Capacities," Chomsky claims that if "we establish the reliability of the judgments and give substantial independent evidence for the theoretical constructions, showing that the postulated principles explain many other facts of a similar nature, withstand empirical test in English and other languages etc." (p.11), this is sufficient grounds for claiming that the grammar is psychologically real. He rejects the claim that psychological testing is required before we can attribute "psychological reality" to our hypothetical grammatical constructs.[17] Chomsky maintains that his views correpond to standard scientific practice which he illustrates with the following account of how astronomers normally conduct their scientific inquiries:

> Consider the problem of determining the nature of the thermonuclear reactions that take place in the interior of the sun. Suppose that available technique only permits astronomers to study the light emitted at the outermost layers of the sun. On the basis of the information thereby attained, they construct a theory of the hidden thermonuclear reactions, postulating that light elements are fused into heavier ones, converting mass into energy, thus producing the sun's heat. Suppose that an astronomer presents such a theory, citing the evidence that supports it. Suppose now that someone were to approach this astronomer with the following contention: True, you have presented a theory that explains the available evidence, but how do you know that the constructions of your theory have physical reality--in short, how do you know that your theory is true? The

> astronomer could only respond by repeating what he has already presented. Here is the evidence available and here is the theory that I offer to explain it. The evidence derives from investigation of light emitted at the periphery. We might want to place a laboratory inside the sun to obtain more direct evidence, but being unable to do so, we must test and confirm our theory indirectly. One might argue that the evidence is inconclusive or that the theory is objectionable on some physical (or, conceivably, methodological) grounds. But it is senseless to ask for some other kind of justification for attributing 'physical reality' to the constructions of the theory, apart from considerations of their adequacy in explaining the evidence and their conformity to the body of natural science, as currently understood. There can be no other grounds for attributing physical reality to the scientist's constructions [Chomsky 1976a pp.4-5].

Chomsky admits that new and more direct methods of studying events taking place inside the sun may be devised but denies that any empirical evidence will ever be totally conclusive. The skeptics will still be in a position to insist on additional evidence before they will accept the astronomer's claim of physical reality. This infinite regression is also possible in linguistics:

> . . . like the astronomer dissatisfied with study of light emissions from the periphery of the sun, we can search for more conclusive evidence, always aware that in empirical inquiry we can at best support a theory against substantive alternatives and empirical challenge, not prove it to be true. It would be quite reasonable to argue against a claim for psychological reality--i.e., truth of a certain theory--on the grounds that the evidence is weak and susceptible to explanation in different terms; needless to say, the evidence that supports the linguist's constructions is incomparably less satisfying than that available to the physicist [Chomsky 1976 pp.5-6].

One can readily agree with Chomsky that an astronomer who has considered all the available evidence and is not confronted with a set of competing theories (all in accord with the facts) is free to place a 'realist interpretation' on his theory. One cannot agree, however, that this is the situation that obtains at the moment in linguistics. One of the more critical problems confronting the linguist at this time is the fact that the standard categories of linguistic data grossly

underdetermine linguistic hypotheses. For every set of linguistic data there is a superabundance of grammars that accurately predict the data. As discussed in Chapter 3, there is no principled basis for selecting one of these grammars over all the others as the "psychologically real" grammar.

It is also not the case that the linguist, like the astronomer, has considered all available categories of data. It may be the case that the diehard skeptic can always insist on additional evidence before he/she will accept a "realist interpretation," but this does not excuse the linguist from the responsibility of considering all available forms of data relevant to a claim before advancing the claim that the hypothesis is true of the real world. It seems to me that it is the duty of the scientist to make a sincere effort to consider all relevant categories of data. If we don't insist that scientists have this duty, scientists will be free to consider whatever subpart of the data they find convenient. It is unfortunately the case thatmany scientists would find it convenient not to consider any evidence which places their theory in jeopardy. It seems foolhardy to permit an astronomer to advance a claim about the nature of the sun solely on the basis of indirect evidence if there is more direct evidence available or to permit a geologist to advance claims about the geomorphology of Mars solely on the basis of evidence available from earthbound observations when the Mariner probe provides us with new and more direct data. If we insist that scientists confront as much of the available data as is possible, we can cut down on the number of false theories which are proposed. Science does not advance very quickly if scientists waste their time refuting theories that should never have been advanced in the first place.

The primary reason the scientist cannot ignore whatever evidence he pleases to, of course, is that the evidence he chooses to ignore may contain counterexamples to the theory that is is proposing. Presumably, scientists are interested not only in making linguistic theories testable but also in testing them. There is just no scientific justification for attempting to ignore new types of potentially disconfirming evidence. No matter what discipline the evidence is exiled to, there is still a general scientific requirement of goodness of fit between related scientific domains. Chomsky must therefore confront the negative results from both psycholinguistics and automata theory and either explain these results away or abandon his claim that transformational grammars are psychologically real. Meanwhile, the rest of us can rejoice in the prospect that evidence from psycholinguistics, etc., may be able to help us reduce our superabundance of grammatical hypotheses to a more manageable level.

1.4 Conclusions

In this chapter we have been evaluating whether or not linguistics should be a mentalistic science and whether or not transformational grammars of the types that have been proposed in the literature are in any sense "psychologically real" (i.e., in any sense part of an adequate mentalist general linguistic theory). Chomsky's claim that linguistics should be a mentalistic science whose goal is to explain linguistic competence and its acquisition has recently been challenged by a number of critics (1) on the grounds that the proper subject matter for linguistic investigation is language, not knowledge of language, (2) on the grounds that Chomsky's enlarged domain of inquiry has yet to be shown more efficient, productive or natural than the traditional domain of linguistic

inquiry, (3) on the grounds that the mentalist (or conceptualist) approach makes counter-intuitive claims about what constitutes a speaker of a natural language, (4) on the grounds that the mentalists falsely claim that grammatical rules are subject to falsification via psychological or neurological testing when in fact grammatical rules are no more subject to psychological disconfirmation than are mathematical truths, (5) on the grounds that linguistics should be concerned with axiomatizations while psycholinguistics should be concerned with algorithms for language use, and (6) on the grounds that mentalist restrictions on the form of grammars force us to provide nonoptimal, "untrue" accounts of human languages.

It has been argued here that there is no reason that the linguist cannot be concerned with both language and knowledge of language, since no conceptual confusion necessarily ensues from dealing with both topics. In order to deal adequately with the question of what is the precise nature of human linguistic competence, researchers will have to be familiar with the structure of language as well as techniques for determining the precise contents of human knowledge. There is no reason why it should be the psychologist and not the linguist who investigates this enlarged domain. The mentalist approach does not entail any counter-intuitive claims about what constitutes a speaker of a language, and the mentalist does not claim that psychlinguistic or neurolinguistic tests can disprove grammatical rules. What the mentalist claims is that psycholinguistic and neurolinguistic tests can disprove claims that these rules are part of the grammar internalized and used by speakers of the language. The rules in question may very well be part of another grammar

of the language which is not internalized by speakers of the language. The distinction drawn between axiomatizations and algorithms on the basis of "revealingness" is contrived and false. In any case, no one has demonstrated that speakers manipulate their language with a set of unrevealing rules of any type. Nor has anyone demonstrated that we need any additional loosening of grammatical restrictions to come up with optional "true" grammars. If anything, the evidence shows that we need more restrictive theories. It is unlikely that any autonomist is going to make much headway with the claim that mentalists do not have the optimal "true" grammar of the language until an acceptable definition of "optimal" is provided along with evidence that suggests that only a single grammar will meet its criteria and that that grammar is outside the class of possible human grammars. It's also not clear how one of the set of grammars that accurately predict the linguistic data is the "true" grammar while other grammars are "false." It is not self-evident that the notion of truth applies here at all. All we seem to need is the distinction is/is not a grammar of the language. Since none of the criticisms of the mentalist approach has been valid, it seems reasonable to conclude that the linguist is free to develop a mentalist linguistics. The research goals of the mentalist are unquestionably worthwhile.

Agreement with Chomsky's goals is, of course, not agreement with Chomsky's claim that he has made significant progress toward attaining those goals. Psycholinguistic tests reveal that no version of transformational grammar tested to date plays an active role in language use. Results from automata theory reveal that no model of the speaker-hearer's competence that incorporates a grammar functioning as a random generating

device (whatever the power of that grammar may be) will be an adequate model of the speaker-hearer's competence because there is no way of accessing the output of that device in real time. Analysis of alternative roles for transformational grammars to play in connection with linguistic competence reveals that none of these roles is viable. It is the conclusion of this chapter that linguists must turn their attention to developing a functioning performance model. This model cannot be a loose patchwork of "heuristics" but rather must be a full-fledged algorithm. There is no reason to believe that the linguist will have to provide anything more than this to meet the mentalist goals. As we noted in this chapter, this "performance" model is as abstract and idealized as anything Chomsky has ever proposed as a "competence" model and it doesn't seem to require supplementation by any other sort of grammar. Studies of learnability by automata theorists and the linguist's own best estimates of how much power our grammars will need suggest that the power of the human grammatical systems must be no greater than that of a primitive recursive automata.

Footnotes

[1]The reader should not be misled into thinking that LSLT is totally devoid of references to speakers' linguistic abilities (see pp. 113-114). On the other hand, the reader should also be warned that there are quite a few scholars who will take Chomsky's claims of an implicit "psychological analogue" with a grain of salt because of some extremely strong anti-mentalist sentiments that were expressed in LSLT (some of which were toned down for the 1975 edition).

[2]Despite the use of the term "competence" in this passage, the terms "competence" and "performance" are not used in either LSLT or Syntactic Structures. Chomsky maintains, however, that the distinction is clear throughout (see p. 7 of the 1975 introduction to LSLT).

[3]At the point when this paper was written, Chomsky's linguistic theory did not include a semantic component, so the closest he could come to providing a semantic interpretation was to recover the underlying structural description.

[4]The discussion section of "The Logical Basis of Linguistic Theory" (1964b) contains a lucid explanation of the difference between the recognition procedure and the grammar (cf. p. 996).

[5]This warning was reissued many times. Perhaps the most important subsequent discussion of the difference between the grammar and the production and recognition devices it is incorporated into appears in "Deep Structure, Surface Structure and Semantic Interpretation" (1970) where Chomsky discusses the proper interpretation of directionality (see pp. 70-71).

[6]Chomsky, on occasion, speculates that transformations may serve to get around memory limitations (see p. 156 of "The Formal Nature of Language" [1967]).

[7]Chomsky periodically claims that there should not be separate grammars for production and perception (see, for example, p. 45 of Chomsky's 1964c discussion of Miller and Ervin).

[8]Chomsky (1958) indicates that the child internalizes a grammar:

> ". . . on the basis of observations of a certain set of sentences and non-sentences (corrections by the verbal community)--it is generally assumed that these play an important role in language learning, but there is little evidence for this, and no indication that the task of constructing a device of the type (4) is simplified if they are admitted as available inputs" (p.431).

[9]The autonomists include Fred Dretske, Esa Itkonen, Michael Kac, Jerrold Katz, Gerald Sanders, and Scott Soames.

[10]For a particularly clear statement of Chomsky's views on why linguistics should be a mentalistic science see pp. 3-5 of Reflections on Language (1975).

[11]See Minsky (1967) on the relation between axiomatizations and algorithms.

[12]It should be noted that the point of this example was to question whether we would ever have a sufficient number of valid criteria to select only one of the set of devices which make correct predictions about the facts as the optimal account of those facts as the "optimal" device. I am obviously not claiming that the Platonist account of time would be in the form of a physical clock. It might be worth noting, however, that, in the case of a Platonist theory of language, the Platonist must select one of the set of possible grammatical descriptions as the optimal or true theory of the language. The grammars in the set of possible human grammars are all abstract objects. The Platonist cannot have an abstract grammatical account of the language that is not a member of the set of abstract possible grammatical descriptions. Any abstract grammatical account of the language, including the Platonist's, must by definition be a member of the set of abstract grammatical descriptions.

[13]A set is "strictly recursively enumerable" if it is recursively enumerable but not recursive. A language is recursive if both it and its complement (the set of strings that are not part of the language) are recursively enumerable. A set is recursively enumerable if there is an effective procedure (a procedure guaranteed to halt in a finite number of steps) for enumerating each of the members of the set.

[14]Cf. E. Gold (1967) "Language Identification in the Limit."

[15]It is difficult to see how Chomsky can make claims about the "psychological reality" of his approach (which contains components whose contents and manner of operation are unknown) and at the same time feel free to accuse Skinner of dogmatism.

[16]Gold's 1967 results are not widely known or appreciated by linguists. I learned of the existence of these results from Constantine Kaniklidis, who is a cornucopia of information about obscure but interesting results in automata theory.

[17]Chomsky points out (p. 6) that

> "The literature takes a rather different view. Certain types of evidence are held to relate to psychological reality, specifically, evidence deriving from studies of reaction time, recognition, recall, etc. Other kinds of evidence are held to be of an entirely different nature, specifically evidence deriving from informant judgments as to what sentences mean, whether they are well formed, and so on. Theoretical explanations advanced to explain evidence of the latter sort, it is commonly argued, have no claim to psychological reality, no

matter how far-reaching, extensive, or persuasive the explanations may be, and no matter how firmly founded the observations offered as evidence, to merit the attribution of 'psychological reality,' the entities, rules, processes, components, etc., postulated in these explanatory theories must be confronted with evidence of the former category" (Chomsky 1976a p.6).

Chomsky, however, argued on page 12 that

"Some evidence may bear on process models that incorporate a characterization of linguistic competence, while other evidence seems to bear on competence more directly, in abstraction from conditions of language use. And, of course, one can try to use data in other ways. But just as a body of data does not come bearing its explanation on its sleeve, so it does not come marked 'for confirming theories' or 'for establishing reality'" (Chomsky 1976a p.12).

Chomsky is correct in claiming that all the data has a bearing on the question of psychological reality. He is not, however, correct in assuming that this fact rescues him from all possible objections to his approach. There are numerous competing grammatical descriptions for each language that accurately predict the data. To select one of these and claim that it, rather than the others, is psychologically real, we need supporting evidence from psycholinguistics. If there was only one grammatical description in the set of possible grammatical descriptions that was compatible with the data, we would be free to advance a claim of psychological reality without further psycholinguistic evidence--although it seems reasonable to assume that, even under these circumstances, the scientist would be pleased to have a second check on the validity of his argumentation and the exhaustiveness of his seach and would, therefore, want to examine the available psycholinguistic evidence.

CHAPTER 2

ON SPECIFYING THE SET OF LINGUISTIC UNIVERSALS

The Language Acquisition Device (L.A.D.) is usually divided into two major subcomponents: (1) the set of linguistic universals and (2) the evaluation measure. This chapter is concerned with the set of linguistic universals which define the notion, possible human grammar. We will be seeking to determine whether or not the linguist has sufficient empirical evidence to be able to determine with some significant measure of certainty the precise set of linguistic universals which correctly characterizes the set of possible human grammars. Chomsky, as we noted in Chapter 1, maintains that the claims he makes about the existence of specific linguistic universals are empirical in nature. The question of whether or not the existence of an innately specified set of linguistic universals can help to explain how children acquire language is deferred to Chapter 4 (where we examine the adequacy of the entire Language Acquisition Device as an explanatory model of the acquisition of language.) Our analysis of this chapter begins with a presentation of Chomsky's view on universal grammar and then examines the testability of claims about the existence of specific linguistic universals.

2.1 Chomsky's Conception of Universal Grammar

The set of linguistic universals defines such notions as possible sentence, possible structural description and possible human grammar and, in addition, provides us with algorithms for enumerating the

set of possible sentences, the set of possible structural descriptions and the set of possible human grammars. Linguistic universals are usually divided into two major types--formal universals and substantive universals. Formal universals specify the form of the linguistic rules contained in the various subcomponents of the grammar (e.g., the rules may be said to be structure preserving, or obey the A over A constraint, or not permit certain kinds of deletions, etc.). Substantive universals specify the possible content of linguistic rules--the theoretical vocabulary that can be drawn upon (e.g., the phonetic features, syntactic categories, semantic markers, etc., that are referred to in linguistic rules). Some scholars distinguish a third type of linguistic universal (usually called an organizational universal) which describes the subcomponents of the grammar and their interaction and within the subcomponents describes the interaction of the rules (e.g., the rules might be said to apply cyclically, etc.). Some universals may be universals in the strong sense that every language contains or conforms to them; or they are universals in the weaker sense that every language may, but need not, utilize them (e.g., languages are free to utilize any of the phonetic features but no language exists that makes distinctive use of all of them). The goal of the linguist is to uncover the "true universal"--which Chomsky defines in his 1977 paper "Filters and Control" as follows:

> By a 'true universal' we mean a principle that holds as a matter of biological necessity and therefore belongs to UG, as contrasted with a principle that holds generally as a matter of historical accident in attested languages. The distinction may be difficult to establish, but it is fundamental [Chomsky 1977 p.437].

While Chomsky may not have always claimed that the general linguistic theory can do double duty as a theory of language acquisition,

he has always claimed that the principles of general linguistic theory are empirical hypotheses that are all too easily disconfirmed by linguistic data. Chomsky claimed in The Logical Structure of Linguistic Theory (1955/75) that the fact that certain clear linguistic cases must be adequately and simply described across numerous languages places a very heavy empirical check on the general linguistic theory:

> The general theory must meet the condition that all grammars to which it leads must satisfy whatever external criteria of adequacy we can establish. . . . It is important to recognize that even weak conditions of adequacy may impose severe restrictions on the choice of grammars for a given language L_1. While many grammars of L_1 may meet these conditions when L_1 is considered in isolation, it may be the case that very few of these grammars follow from some general theory that leads to grammars of the languages L_2, L_3, ... all of which meet these conditions of adequacy [Chomsky 1955/75 p.81].[1]

The external criteria of adequacy that the grammars defined by the general linguistic theory must meet are, of course, the linguistic intuitions of native speakers. The predictions these grammars make must correspond to the speakers' intuitions of grammaticality, ambiguity, synonymy, etc.:

> . . . we pointed out that if linguistic research is to have content, there must be certain criteria of adequacy for each grammar, outside of those internal to linguistic theory. But we have said very little about what these criteria may be. Actually, there seems to be very little to say about this that is not uncomfortably vague. Clearly the sequence generated by the grammar as grammatical sentences must be acceptable, in some sense, to the native speaker, and the processes described must conform somehow to his 'habits.' Above . . . we construed the linguist's task as one of reconstructing in some systematic and inspectable way the speaker's 'linguistic intuition.' . . . The speaker's intuition about form . . . poses, for each language, conditions that must be met by linguistic theory [Chomsky 1955/75 p.101].[2]

In addition to providing grammars which accurately predict speakers' linguistic intuitions, the grammars provided by the theory must be simple and revealing.

The question naturally arises as to the best research strategy for investigating linguistic universals. In his 1966 paper "Linguistic Theory," Chomsky points out that there are two distinct approaches to uncovering linguistic universals:

> One way is by an investigation of a wide range languages. Any hypothesis as to the nature of linguistic universals must meet the empirical condition that it is not falsified by any natural language. . . . But there is also another and, for the time being, somewhat more promising way of studying the problem of universals. This is by deep investigation of a particular language, investigation directed towards establishing underlying principles of organization of great abstractness in this language [Chomsky 1966 pp.57-58].

Chomsky maintains that when we investigate linguistic universals by conducting a deep analysis of a single language, one particularly important clue that a principle used in a linguistic description is universal rather than language-specific is the difficulty we would have in explaining how such a principle could have been learned by the child. If it seems unlearnable, we hypothesize that the principle is an innately specified linguistic universal. This argument form is schematized in Chomsky's 1976 paper "On the Nature of Language":

> Suppose we find that a particular language has the property P: that is, speakers' judgments and other behavior conform to P where clear and reliable, constructions violating P are rejected, and so on. Suppose, furthermore, that P is sufficiently abstract and evidence bearing on its sufficiently sparse and contrived so that it is implausible to suppose that all speakers, or perhaps any speakers, might have been trained or taught to observe P or might have constructed grammars satisfying P by induction from experience. Then it is plausible to postulate that P is a property of the language faculty, that languages conform to P as a matter of biological necessity [Chomsky 1976b p.47].

The classic examples of unlearnable principles are the cycle in syntax and phonology and the principle that linguistic rules are structure dependent. If a principle is innately specified it, of course, doesn't have to be learned. Chomsky indicates that when we postulate that the

general linguistic theory is also a model of the innately specified language acquisition device, we are claiming that the child cannot learn any grammar or language that doesn't correspond to the innately specified form in the manner in which language is normally learned--although we might be able to learn a nonhuman grammar or language in the same way that we can learn any scientific theory. The tests that are suggested by this claim cannot, of course, be conducted using children.

Chomsky suggests that it is desirable to postulate as many linguistic universals as possible--not just because in so doing we shed new light on what it means for something to be a human language or grammar and shed new light on human cognitive abilities, but because he assumes that the fewer potential grammatical hypotheses that the child has available (i.e., the fewer grammars there are in the set of possible human grammars), the more plausible a model of language acquisition which requires the child to search through this set of possible human grammars in order to find the correct grammar becomes. Any universal which is compatible with the linguistic evidence is to be embraced as a partial solution to the problem of whittling down the class of possible human grammars to manageable proportions. Chomsky also maintains that the more restrictive we make the specification of the class of possible human grammars the more falsifiable it becomes.

2.2 The Testability of Hypotheses Concerning Linguistic Universals

As noted in Chapter 1, the goal of the mentalist linguist is to narrow the set of all possible grammatical descriptions of human languages to the set of all humanly possible grammatical descriptions of possible human languages. In the 1950s, Chomsky assumed without evidence

that while there might be a superabundance of grammars from a bewildering variety of linguistic models which are capable of accurately describing any one particular natural language (i.e., capable of accurately predicting speakers' intuitions of grammaticality, ambiguity, synonymy, etc.), the fact that such general linguistic theorieswere required to provide grammars of comparable descriptive accuracy for the entire range of possible human languages would place such a severe empirical constraint on the general linguistic theories that we will be able to select a single general linguistic theory.

It is clear that the condition that descriptive accuracy must be achieved for all languages is strong enough to rule out numerous linguistic universals and general linguistic theories (i.e., it is obviously going to be possible in principle to devise general linguistic theories which partition the set of possible grammatical descriptions in such a fashion that the proposed set of possible human grammars does not contain an accurate grammatical description for some particular natural language). There have, in fact, been a number of proposed linguistic universals that have been convincingly refuted when they were subjected to cross-language testing. But a demonstration that some linguistic universals will be vulnerable to empirical disconfirmation is not a demonstration that they are all empirically testable (or even a demonstration that the majority of them will be falsifiable). The evidence suggests that the external conditions of adequacy are a much weaker empirical constraint than Chomsky imagined. There seem to be numerous universals that are not vulnerable to falsification on the basis of standard linguistic data. The paradigmatic case of this sort is, of course, the Universal Base Hypothesis.

In the late 1960s some linguists suggested the possibility that there might be a single universal base component for all human grammars. The existence of a universal base component would ease the task of learning language since there would be one less grammatical component for the child to learn. It was thus quite disturbing when the strong claim that speakers have the same universal base component (from which the surface structures for all natural languages are transformationally derived) was shown to be unfalsifiable. This result was proven in Peters and Ritchie's 1969 paper, "A Note on the Universal Base Hypothesis" and Peters' 1970 paper, "Why There are Many 'Universal Bases.'" Peters and Ritchie show that, given the power of standard transformational grammars (which are equivalent in power to unrestricted rewriting systems or Turing machines), it is a foregone conclusion that universal bases exist from which we will be able to derive accurately the correct predictions of the speakers' grammaticality, ambiguity, and paraphrase judgments, and this will be the case even if we reduce the power of the base to that of a context-free grammar or even if we reduce the power of the base still further to that of a finite state grammar. Given the power of transformations to alter the structures provided by the base component, there will, in fact, be an infinite number of universal bases from which we can successfully derive the correct surface structures for every natural language. There is, then, no evidence internal to linguistics that would disconfirm the hypothesis that speakers have a single universal base component and there is no evidence that would lead us to prefer one universal base over its infinite number of brethren.

Chomsky, of course, never claimed that there is a universal base

component. But this does not mean that Peters and Ritchie's results do not apply to the claims that Chomsky has made about the form of the base component. If anything, it would be more difficult to falsify a theory that just states the form of the rules in the component and not their substance. It is guaranteed that whatever Chomsky has said or will say about the form of the base component will not be falsifiable as long as the transformational component has the power of an unrestricted rewriting system. If the transformational component has the power of an unrestricted rewriting system there are guaranteed to be base components of the proper form for every language from which some transformational component(s) can generate the proper predictions.

The significance of Peters and Ritchie's results is not limited to questions concerning the form of the base component. There are, for instance, an astronomical number of distinct general linguistic theories which are equivalent in power to an unrestricted rewriting system. Since unrestricted rewriting systems can output (predict) the proper results for any system that is well defined (i.e., denumerable), it follows that our standard linguistic intuitions will not provide any evidence for choosing among general linguistic theories of such power. Native speakers (and linguists) unfortunately have no valid intuitions concerning the form of the grammars they have internalized, only intuitions about its predictions. The existence of multiple general linguistic theories that are compatible with the data is, of course, not entirely a byproduct of the excessive power of these models. The problem is that at every level of power there will be numerous alternative grammatical devices which are capable of generating the same sets of possible

human languages. There will thus be numerous general linguistic theories incorporating these different devices which cannot be empirically distinguished on the basis of the standard linguistic data. If we wish to be able to determine which of these general linguistic theories corresponds to the child's language acquisition device, we must bring data to bear from psycholinguistics. Chomsky's claim that we can dispense with such data is, as we noted in Chapter 1, in error. The linguist's ordinary range of data grossly underdetermines the selection of a specific linguistic theory and specific linguistic universals.

Adding to the linguist's problems with "testability" is the fact that linguistic theory provides the linguist with numerous devices which can serve to protect false hypotheses. Even if these devices can be independently motivated, they nonetheless provide the linguist with the undesirable power to save hypotheses which should be discarded. Obvious examples of such devices are: (1) the distinction between [illegible] and a counterexample and (2) the distinction between [illegible] acceptable but ungrammatical and strings that are [illegible] grammatical. The first distinction permits the linguist [illegible] that a counterexample is just an exception to be handled by labeling the offending linguistic unit with an exception feature and the second device permits the linguist to claim that some counterexample a critic brought up is actually ungrammatical and need not be accounted for despite its acceptability or that some case of overgeneration is not a case of overgeneration but rather a case of an unacceptable but grammatical sentence. The problem arises from the fact that there are insufficient constraints on the linguist's use of these distinctions. The claim, for example,

that some acceptable sentence is ungrammatical should flow from an independently motivated theory of the psychological factors that can interfere with judgments of grammaticality. Without such a constraint, a counterexample to a proposed universal can be ignored for the small price of a claim that when we do have an independently motivated theory of the psychological factors that interfere with linguistic judgments, this theory will explain why speakers find ungrammatical strings like the suggested counterexample acceptable. It is possible to provide independent evidence for claims of this sort (as Langendoen and Bever have shown in several papers). We must insist that such evidence be provided.

Indeterminacy and unfalsifiability are not the only problems with Chomsky's approach to universal grammar. Let's start with Chomsky's methodological claim that if we cannot explain how some feature or principle postulated in our grammatical description can possible be learned, we should further postulate that the feature or principle in question is a part of universal grammar. If we know for a fact that speakers have internalized the grammar we are postulating, and we know that no learning theory is in principle capable of learning certain principles embodied in the grammar, and we know that a grammar of this sort can be acquired by a language acquisition device of the type Chomsky is envisioning if the requisite principles are built into the innate specification of the form of the grammar, then we can reasonably claim that the principle(s) in question are part of universal grammar. But we don't know what grammar the child has acquired (that is the goal, not the given, of our study). Our knowledge of learning procedures is not at all highly developed. And, as I argue in Chapter 4, no model of language acquisition

which involves a search through the set of possible human grammars can explain actual language acquisition since the procedure, in principle, cannot operate in real time. To decide that some principle is a linguistic universal, we must have evidence that the grammars speakers are capable of internalizing actually conform to this principle. This evidence must come from psycholinguistics since linguistics is confronted with a multiplicity of general linguistic theories which embody different universal principles and yet are all capable of accurately describing the entire range of linguistic facts.

Another problematical assumption that Chomsky makes is the assumption that placing additional constraints on the form of grammars automatically narrows down the size of the set of possible human grammars from which the child gets a grammatical description and, therefore, makes the task of language learning more reasonable. Paradoxical as it may seem, these constraints do not decrease the size of the set of possible human grammars. Many intuitively satisfying assumptions fall by the wayside when one deals with infinite sets. The set of natural numbers (i.e., the set of positive whole integers) is the same size as the set of positive even numbers and the same size as the set of positive odd integers. All three sets are of cardinality 1 (the first order of infinity). The linguistic constraints that have been proposed narrow down the set of possible human grammars from a denumerably infinite set of the first order of infinity to a denumerably infinite set of precisely the same cardinality. In both cases the child is confronted with a denumerably infinite set of grammars to search through. A linguistic constraint would, however, make a significant difference if it narrowed

down the class of possible human grammars from a nondenumerably infinite set to a denumerably infinite set or a denumerably infinite set to a finite set.

Chomsky also claims that when we constrain our specification of the possible form of grammars, we not only ease the task of language learning but also make our theory more falsifiable. This view was clearly expressed in some of Chomsky's comments on Postal's theory:

> Postal argues that the more homogeneous theory that he proposes is a 'better theory', on the grounds that it contains fewer devices, is simpler, and so on. I argue that it is a 'worse theory' on the grounds that it is less falsifiable and permits a far wider range of grammars than the alternatives considered. . . . Given two theories T and T', we will be concerned not merely with their simplicity and homogeneity, but also with the restrictiveness. If T and T' both meet the condition of descriptive adequacy, but T permits only a proper subset of the grammars permitted by T', then we may well prefer T to T', even if it is more complex, less homogeneous. . . . Postal's homogeneous theory H is, as he says, simpler than the 'extended' standard theory, ST. It is also less restrictive and less falsifiable, and thus less capable of dealing with the projection problem and the issues raised at the level of explanatory adequacy [Chomsky 1974 pp. 47-48].

It is not clear, however, why Chomsky feels that the claims his theory makes are more restrictive or falsifiable than those made by Postal's theory (or any other theory capable of describing human languages). Both theories divide the set of possible grammatical descriptions of possible human languages into two sets: (1) the set of possible human grammars and (2) a set of grammars which describe human languages but are not possible human grammars. Since both theories make a prediction about which set each grammar in the set of possible grammatical descriptions belongs to, they make the same number of predictions. They differ, however, on the predictions they make as to the contents of the two sets. In order for Chomsky's theory to be more falsifiable than

Postal's there must be some prediction it makes that can be tested while the equivalent prediction from Postal's theory cannot. But there are only two choices we can make--a grammar is either in the set of possible human grammars or it is not. It is, therefore, impossible to make a determination of the correctness of a prediction made by Chomsky's theory about some grammar without simultaneously ascertaining the correctness of the prediction Postal's theory makes about that grammar. All general linguistic theories that make precise claims will be equally falsifiable in this sense. Any claim about the status of some grammar that is testable for one will be testable for them all.

The last problem with Chomsky's approach to linguistic universals that will be raised here concerns a methodological principle that Chomsky invokes when debating with other linguists about the status of proposed universals. Ordinarily, the scientist who is confronted with a set of competing hypotheses of comparable simplicity and predictive accuracy has no principled justification for designating one of these hypotheses as the correct or true hypothesis concerning the data. If, however, there is only one theory under consideration which accounts for the facts, the scientist would accept that theory and many scientists would place what Chomsky calls a "realist interpretation" on the theory (i.e., they would claim that the theory is in some sense true of the world). Now, what happens when we subsequently develop an alternative theory which is of comparable simplicity and predictive power to our original theory? Chomsky claims that we should continue to maintain that the first theory proposed is the correct or true one until it is actually refuted.

If this principle is accepted, then the first linguist to propose a linguistic universal capable of explaining some set of facts stands an extremely good chance of having his/her claim viewed as the truth for the foreseeable future given the invulnerability of most of these claims to disconfirmation from the linguistic data. The claim would stand forever if we accept Chomsky's self-serving arguments against the use of data from psycholinguistics. But what justification is there for conflating historical precedence with logical or scientific precedence? The historical accident that a theory or hypothesis was advanced at a time when no other alternative theories were available supplies us with no grounds for designating one of a set of equally simple and predictively accurate descriptions as the true description. Linguistics is not a card game where the house has the edge and must be beaten in order to be dethroned. To select a hypothesis on the basis of the temporal order in which it was proposed is to invoke a criterion that is as irrelevant as selecting a theory on the basis of the age, nationality, sex, religion, etc., of its author. Recent claims by Chomsky and others that a counterexample is uninteresting unless accompanied by an adequate reanalysis of the facts represent a further weakening of the falsifiability of linguistic theories. It should be rejected. If a theory has been refuted what conceivable justification can there be for continuing to work within such a framework? It may be a sociological fact that scientists will continue to work in a refuted paradigm until a new one comes along, but this is a regrettable sociological fact, not a valid methodological principle.

2.3 Conclusions

In this chapter we have been attempting to assess what success linguists have had in "homing in" on a specification of the form of possible human grammars. We have discovered that Chomsky's assumption that the external criteria of adequacy (the speaker-hearers' linguistic intuitions) would provide a strong test of the adequacy of our general linguistic theory has proven false. Any linguistic theory that has the power of a Turing machine provides grammars that are capable of predicting any well-defined set of linguistic data. There are a staggering number of general linguistic theories of this power (including Chomsky's "Standard Theory"). At every level of power, if there is one general linguistic theory that is capable of providing grammars which account for the facts of language there will be many. It is not particularly surprising, then, that linguists have not been able to settle on a single general linguistic theory of the possible form of human grammars solely on the basis of the linguistic evidence. I can't see any way that we will ever be able to settle conclusively on a single general linguistic theory without conclusive evidence from psycholinguistics concerning the principles that the child has available for acquiring a language. Since no one seriously anticipates conclusive evidence from psycholinguistics on this matter for the foreseeable future, the linguist will have to become inured to the fact that there will be numerous competing general linguistic theories that cannot be distinguished empirically. I don't feel, however, that we should become inured to any of the present set of competing theories (standard theory, trace theory, case grammar, generative semantics, relational grammar, Montague grammar,

dependency theory, etc., etc., etc.) that have been shown to be equivalent in power to unrestricted rewriting systems. If there is one thing that the learnability studies of automata theorists have demonstrated, it is that if we want our grammars to be learnable they should be no greater in power than primitive recursive automata.

Linguistics is, of course, not unique in having an indeterminacy problem. The theories of every empirical science are underdetermined by the data. It would have been very surprising, then, if the linguist had come up with something definitive to say about the speaker's linguistic competence and its acquisition (especially after only two decades of study). The proper conclusion to draw from the fact that linguistics is afflicted with a serious indeterminacy problem is not that linguistics should be abandoned as a scientific pursuit but rather that linguists should refrain from making premature claims about the "psychological reality" of their linguistic theories and concentrate their attention on finding additional sources of evidence that would help reduce the indeterminacy problem.

If linguists are to have any serious hope of learning anything definite about the possible form of human grammars, they must make every effort to guarantee that their theories are both testable and tested. There are no categories of data or types of test we can afford to eschew if we wish to reduce our indeterminacy problem and make a genuine contribution to our understanding of the human mind and its capacity for language learning and use. Nor can we afford to take advantage of the hypothesis-protecting devices our theory makes available or afford to utilize the fallacious arguments that were uncovered in this chapter

that linguists have used to falsely inflate the status of their pet theories. Linguists must also show a greater commitment to reducing the power of general linguistic theories than they have thus far demonstrated. General linguistic theories seem to be proliferating without much consideration on the part of linguists of the need to provide theories that are not equivalent in power to unrestricted rewriting systems.

In addition to assessing whether or not linguists are in a position to state anything definitive about the possible form of human grammars that is not hopelessly abstract and insubstantial, we have been trying to assess what kind of contribution linguistic constraints make toward explaining language acquisition. Chomsky has repeatedly stated that linguistic universals play a crucial role in explaining language acquisition in that they narrow down the set of possible linguistic hypotheses that the child must consider. It is not clear, however, that the linguistic universals ease the burden of language acquisition, since they narrow down a denumerably infinite set (cardinality 1) to a denumerably infinite set (cardinality 1). None of the constraints that have been proposed in the literature reduces the size of the set of possible human grammars below the first order of infinity and most linguists do not, in fact, want to claim that there are only a finite number of possible human grammars.

Footnotes

[1]See p. 14 of Syntactic Structures for another statement of this claim.

[2]Chomsky was, however, a bit uneasy at the time about using "intuitions" as data: "But intuition, of course, is an extremely weak support. The program of linguistic research would be a much clearer one if we could show experimentally that these intuitions have distinct behavioral correlates" (Chomsky 1955/75 p. 101).

CHAPTER 3

THE EVALUATION MEASURE

In this chapter I will examine the history, formulation, motivation, function, and adequacy of Chomsky and Halle's conception of an evaluation measure. The chapter consists of four sections: (1) an historical account of the evolution of the concept of the evaluation measure in Chomsky and Halle's writings, (2) an examination of the motivation Chomsky supplies for adopting the evaluation measure as part of linguistic theory and as part of the language acquisition device, (3) an attempt to delineate what Chomsky and Halle envision the evaluation measure as measuring and an examination of the substantive progress that has been made to date toward providing an accurate, functioning evaluation measure, and (4) conclusions about the merits of past research on the evaluation measure and the proper direction for future research.

3.1 Chomsky's Conception of the Evaluation Measure

In this section Chomsky and Halle's views on the evaluation measure are presented in almost strict historical sequence from The Morphophonemics of Modern Hebrew (1951) through The Sound Pattern of English (1968). This will provide the background essential to any in-depth examination of the evaluation measure. The theory of markedness and subsequent changes in Chomsky and Halle's thinking on the evaluation measure will be presented as they arise naturally in the context of the critical discussion of the evaluation measure and the criticisms that

have been leveled against it. The historical presentation is not totally devoid of criticism--as the reader will notice almost immediately. I have attempted, however, to limit critical comments in this section to taking a stance on issues of textual interpretation and to pointing out certain lacunae in Chomsky and Halle's argumentation.

The goal of devising an evaluation (or "simplicity") metric that can be used to choose between competing grammatical descriptions (grammars) of a language was first enunciated in Chomsky's 1951 Masters thesis, The Morphophonemics of Modern Hebrew. Chomsky's initial conception of the evaluation measure (as presented in his Masters thesis and his 1955 work, The Logical Structure of Linguistic Theory, which contained his dissertation and formed the basis for Syntactic Structures) differed from his later conceptions in several important respects: (1) Chomsky seems to utilize the term "simplicity" in its traditional philosophy of science sense in these works, and (2) he had not yet considered (at least in print) the question of whether or not the evaluation measure played any role in language acquisition. Chomsky was convinced that all grammars must meet requirements of simplicity not for esthetic reasons but because "the motives behind the demand for economy are in many ways the same as those behind the demand that there be a system at all" (Chomsky 1951 p. 1). While considerations of simplicity play a role in the discovery or construction of tentative grammars, Chomsky felt that it was useful to consider the processes of discovery and description separately and concentrate on determining the effect of considerations of simplicity on the formulation of grammars.[1]

From the very beginning Chomsky sought to devise a set of

notational conventions which correlated length of grammatical description with deeper cognitive aspects of simplicity:

> . . . we want the reduction of the number of elements and statements, any generalization, and, to generalize the notion of generalization itself, any similarity in the form of non-identical statements, to increase the total simplicity of the grammar. As a first approximation to the notion of simplicity, we will here consider shortness of grammar as a measure of simplicity, and will use such notations as will permit similar statements to be coalesced. To keep this notion of simplicity from reducing to an absurdity, the notations must be fixed in advance, and must be chosen to be neutral to any particular grammar except with respect to the considerations they are chosen to reflect [Chomsky 1951 p. 5].

In his Masters thesis Chomsky set up a hierarchy of simplicity: the grammar with the fewest number of rules was simplestbut among grammars with an equal number of rules the simplest grammar was judged to be that in which the average length of statement was shortest. The weakest requirement of simplicity called for the "reduction of the length of derivations" (see Chomsky 1951 p. 41).

Chomsky made his first serious attempt to define an evaluation (simplicity) metric in The Logical Structure of Linguistic Theory, which was completed in 1955 (but remained unpublished until 1975). In this work Chomsky elaborates a little further on why it is desirable to try to devise a notational system which converts considerations of simplicity into considerations of length, points out some of the requirements that must be met for a notational definition of simplicity to have any validity, and makes some recommendations as to appropriate notational conventions.

The temptation to devise a notational system which permits us to measure simplicity in terms of notational brevity apparently arises from Chomsky's observation that

> In constructing a grammar, we try to set up elements having regular, similarly patterned, and easily statable distributions, and which are subject to similar variations under similar conditions: in other words elements about which a good deal of generalization is possible and few special restrictions need be stated. It is interesting to note that any simplification along these lines is immediately reflected in the length of the grammar [Chomsky 1955/75 p. 117, emphasis added].

Chomsky is not claiming that all aspects of cognitive simplicity are naturally (or iconically) reflected in the degree of notational brevity[2]--he is merely noting a tendency in linguistics for increases in simplicity (or generality) to be correlated with increases in notational brevity and on the basis of this moderate encouragement is willing to undertake the complex, arduous and possibly futile program of devising a set of notational conventions which can accurately convert all deeper considerations of grammatical simplicity into considerations of notational brevity.

Chomsky realizes that, since it is possible to construct extremely complicated notational schemas which create grammars of almost any shortness we wish (particularly if we are permitted to define notations afresh for every grammar we write), we must, for his approach to have any significance, develop a fixed set of notations and a fixed format for grammars. The simplest grammar will be selected from grammars of the appropriate form and the definition of these notations will constitute the basic part of the definition of simplicity (see Chomsky 1955/75 pp. 66-67, 118). As Chomsky sees it:

> The problem of choosing the correct notations is much like that of evaluating a physical constant. Given criteria of adequacy for grammars of certain languages, we can arrive empirically at notations with the property that the grammars meeting the criteria of adequacy are in fact the shortest, given these notations. In other words, we define simplicity so that in certain cases, the simplest grammars

> are in fact the correct ones. As long as we do not take simplicity to be an absolute ideal, thoroughly understood and specified in advance of theory construction, this procedure is no stranger than attempting to define 'morpheme' in such a way that what we know to be morphemes in some language turn out to be morphemes when we apply the theory to a corpus of utterances in this language [Chomsky 1955/75 p. 118].

The fact that the notations must give the intuitively correct results for the grammars of all possible human languages provides what Chomsky considers a very strong empirical test of the adequacy of the notations.

Chomsky felt that we can provide notational conventions which convert considerations of simplicity into considerations of length by devising notational conventions which permit the coalescence of similar grammatical statements. This favors grammars containing generalizations ("We have a generalization when we can replace a set of statements, each about one element, by a single statement about the whole set of elements. More generally, we have a partial generalization when we have a set of similar (not identical) statements about distinct elements" [Chomsky 1955/75 p. 67]). Once we have devised notations that permit the coalescence of similar statements, Chomsky feels we will be able to determine the amount of generalization present by merely examining the notational length (i.e., the number of symbols) of the grammar ("Other features of simplicity can also be measured in a natural way in terms of length" [Chomsky 1955/75 p. 67]).

There has been considerable controversy concerning Chomsky's use of the term simplicity in the early writings which I have been discussing. The naive reader who encounters these early remarks on simplicity would assume that Chomsky was using the term simplicity as it is generally used and understood in both common parlance and the philosophy of

science. It comes as something of a surprise when the reader learns that in later writings Chomsky and Halle warn the reader not to confuse the "simplicity" metric with a measure of simplicity in its traditional philosophy of science sense. They even go so far as to claim that, while the evaluation measure has been referred to as the "simplicity" or "economy" criterion,

> . . . it has never been proposed or intended that the condition defines 'simplicity' or 'economy' in the very general (and still poorly understood) sense in which these terms usually appear in writings on the philosophy of science. The only claim that is being made here is the purely empirical one that under certain well-defined notational transformations, the number of symbols in a rule is inversely related to the degree of linguistically significant generalization achieved in the rule [Chomsky and Halle 1968 pp. 334-35].

Chomsky and Halle's disclaimer is not quite believable. There are numerous instances in Chomsky's early writings where his use of the term simplicity cannot be interpreted in any way other than its philosophy of science sense.[3] Consider the following statement:

> It is important to recognize that we are not interested in reduction of the length of grammar for its own sake. Our aim is rather to permit just those reductions in length which reflect real simplicity, that is, which will turn simpler grammars (in some partially understood, presystematic sense of this notion) into shorter grammars [Chomsky 1955/75 p. 118, emphasis added].[4]

The language of the two quotations is remarkably similar--the only difference being that in the later passage Chomsky and Halle are denying that they ever wanted to capture the very thing that Chomsky in 1955 claimed we are primarily interested in capturing with the "simplicity" metric. It is also difficult to see how Chomsky, if he truly had a unique conception of simplicity in mind, could have been advising the readers of The Morphophonemics of Modern Hebrew and The Logical Structure of Linguistic Theory to read Goodman and Quine's philosophy of

science accounts of simplicity without warning readers that he was using the term in some entirely different and novel sense and without explicating the difference between the two usages.[5]

Syntactic Structures contains Chomsky's celebrated discussion of the goals of linguistic theory in which he distinguishes three possible goals for linguistic theory: the discovery procedure, the decision procedure, and the evaluation procedure.[6] In this discussion Chomsky claimed that the goal of providing a mechanical discovery procedure for grammars (i.e., "a practical and mechanical method for actually constructing the grammar given a corpus of utterances" [Chomsky 1957 pp. 50-51]) was too strong for linguistic theory. Chomsky's evidence for this claim, however, was not particularly strong. He could, for instance, offer no proof that it was in principle impossible to develop a step-by-step mechanical procedure for constructing the best grammar. He did, however, advance his claim only after conducting a thorough study of the numerous procedures that had been devised in this area to that date.[7] It is important to note that Chomsky rejected the goal of providing a discovery procedure not because it wasn't an important or desirable goal, but because he felt it was not an achievable goal. Chomsky felt that the development of even a partially adequate discovery procedure would be useful in that such a procedure might provide valuable clues for the analyst or lead to a smaller class of grammars to be evaluated (see Chomsky 1957 p. 55).

Chomsky also rejected the "weaker" goal of providing a practical mechanical procedure for determining whether or not a grammar proposed for a given corpus (however it might have been arrived at) is the best

grammar for the language the corpus was drawn from (i.e., the decision procedure) as being too strong a goal for linguistic theory. No evidence was provided in support of this claim.

Chomsky felt that any attempt to provide anything more than an evaluation procedure (a practical and mechanical procedure for determining which of any two proposed grammars for a language is better) would "lead into a maze of more and more elaborate and complex analytic procedures that will fail to provide answers for many important questions about the nature of linguistic structures," while, in contrast, "by lowering our sights to the more modest goal of developing an evaluation procedure for grammars we can focus attention more clearly on really crucial problems of linguistic structures and we can arrive at more satisfying answers to them" (Chomsky 1957 p. 53). But while we may have been forced to lower our sights, Chomsky feels we can take solace in the fact that the goal of providing an evaluation procedure "is still strong enough to guarantee significance for a theory that meets it. There are few areas of science in which one would seriously consider the possibility of developing a general, practical, mechanical method for choosing among several theories, each compatible with the available data" (p. 53). Chomsky provides no evidence to support his claim that the goal of developing a mechanical evaluation procedure is a weaker and more feasible goal than the others. In fact, he gives us no reason to believe that the goal, as he conceives it, is feasible at all. The proof is apparently to be in the pudding--Chomsky informs us that all his claims will be borne out when we actually develop and utilize the evaluation measure.

The actual task of providing a formal definition of simplicity

is not undertaken in Syntactic Structures. The Logical Structure of Linguistic Theory, however, does contain an extended discussion of linguistic simplcity and makes numerous concrete proposals concerning the notational conventions to be used in conjunction with the evaluation measure.[8] Syntactic Structures is also devoid of any reference to the "psychological analogue" of formal linguistic evaluation. Chomsky, in his 1975 introduction to The Logical Structure of Linguistic Theory, claims that, while the "psychological analogue" was not explicitly discussed either in that work or in Syntactic Structures, the possibility of such an analogue lay in the immediate background of his thinking during this period and was in some sense implicit in both works. The psychological analogue of the linguist's methodological problem was explained as follows:

> . . . we have two variants of the fundamental problem of linguistics . . . : under the methodological interpretation, the problem is taken to be the justification of grammars, under the psychological interpretation, the problem is to account for language acquisition. A linguistic theory, under either conception, is explanatory in a proper and appropriate sense. Given such a theory and certain 'boundary conditions' fixed by a corpus of data with a preliminary analysis determined by the application of the primitives of the theory, a grammar is selected. Under the methodological interpretation, it is the speaker-hearer's grammar, chosen by the evaluation procedure from among the potential grammars permitted by the theory and compatible with the data as represented in terms of the preliminary analysis. In either case, the grammar determines a range of linguistic facts relating to the structure of sentences. The speaker-hearer's judgments that the facts are as determined by the grammar are thus explained by the linguistic theory; correspondingly, the theory is refuted if the judgments are not in accord with the predictions of the grammar [Chomsky 1955/75 p. 36].

Chomsky's first published reference to the possibility of a psychological analogue to the linguist's methodological problems seems to occur in his 1958 paper, "Linguistics, Logic, Psychology and Computers." The analogy

wasn't fully worked out, however, until his 1962 paper, "Explanatory Models in Linguistics," appeared.

In Current Issues in Linguistic Theory (1964a), Chomsky described three levels of adequacy that linguistic theories can attain: the observational, descriptive, and explanatory levels of adequacy. Since the evaluation measure is said to play a critical role in determining whether or not the highest level of adequacy (the level of explanatory adequacy) is attained by a linguistic theory, I will describe the levels of adequacy in some detail. Chomsky indicates (on pp. 28-29) that the level of observational adequacy is attained by a grammar if it correctly presents the observed primary data (i.e., if it gives an account of the primary data [the corpus] that is the input to the acquisition device). The level of descriptive adequacy (the second and higher level of success) is attained "when the grammar gives a correct account of the linguistic intuition of the native speaker, and specifies the observed data (in particular) in terms of significant generalizations that express underlying regularities in the language" (p. 28). The grammar is, in other words, concerned with giving a correct account of the output of the language acquisition device. The third, and highest, level of adequacy (the level of explanatory adequacy) is attained "when the associated theory provides a general basis for selecting a grammar that achieves the second level of success over other grammars consistent with the relevant data that do not achieve this level of success." The general basis for selecting a grammar is, of course, the evaluation measure. We may say that "a linguistic theory that aims for explanatory adequacy is concerned with the internal structure of the device . . . , that is, it aims to provide a principled basis, independent of any particular

language, for the selection of the descriptively adequate grammar of each language" (p. 29). Chomsky contends that an explanatorily adequate linguistic theory "suggests an explanation for the linguistic intuition of the native speaker" and "can be interpreted as asserting that data of the observed kind will enable a speaker whose intrinsic capacities are as represented in this general theory to construct for himself a grammar that characterizes exactly this linguistic intuition" (p. 28).

This analysis of the levels of adequacy was refined still further in Aspects of the Theory of Syntax (1965). General linguistic theories can now be spoken of as either meeting or not meeting the level of descriptive adequacy. Prior to Aspects, the only level of adequacy relevant to general linguistic theories was the level of explanatory adequacy. To meet the level of descriptive adequacy a linguistic theory must provide (1) an enumeration of the class S_1, S_2, ... of possible sentences, (2) an enumeration of the class SD_1, SD_2, ... of possible structural descriptions, (3) an enumeration of the class G_1, G_2, ... of possible generative grammars, and (4) a specification of a function F such that SD F(i,j) is the structural description assigned to sentence S_i by grammar G_j for arbitrary i,j.[9] In other words, a theory is descriptively adequate if it makes a descriptively adequate grammar available for each natural language. The general linguistic theory achieves the level of explanatory adequacy by providing, in addition to the four items listed above, (5) a specification of a function m such that m(i) is an integer associated with the grammar G_1 as its value (with lower value being indicated by higher number).[10] In other words, it meets the level of explanatory adequacy by providing an evaluation measure.

The goal of providing an evaluation measure will not be an easy goal to attain. For one thing "a theory may be descriptively adequate, in the sense just defined, and yet provide such a wide range of potential grammars that there is no possibility of discovering a formal property distinguishing the descriptively adequate grammars, in general, from among the mass of grammars compatible with whatever data are available (p. 35). The methodological problem (and the problem for constructing a reasonable acquisition model) is to reduce the range of possible hypotheses (the class of attainable grammars) compatible with given primary data to the point where selection among them can be made by a formal evaluation measure. The way to do this is by adding further structure to the notion "generative grammar" on a principled basis. The principled basis for adding additional structure comes from a determination of the extent to which descriptively adequate grammars have features which can be reduced to universal properties of language: "Real progress in linguistics consists in the discovery that certain features of given languages can be reduced to universal properties of language, and explained in terms of these deeper aspects of linguistic form" (Aspects p. 35). These universals can be incorporated into the definition of "generative grammar" and if the restrictions are significant we will have a reduction in the class of possible hypotheses (grammars). It is conceivable that the class of potential grammars could be so narrowly defined that there will be no more than one grammar compatible with the data for each language and there will be no need for an evaluation measure. But Chomsky considers this extremely unlikely since "all concrete attempts to formulate an empirically adequate linguistic theory

certainly leave ample room for mutually inconsistent grammars, all compatible with primary data of any conceivable sort" (Aspects p. 37). It is also conceivable that we will not be able to make sufficient "principled" restrictions on the class of grammars to be able to construct or choose an evaluation measure on a principled rather than an arbitrary basis. Chomsky provides no evidence for the belief that we can formulate the requisite numbers and kinds of universals needed and all concrete attempts to formulate an empirically adequate linguistic theory to date define too large a class of possible grammars for us to construct or choose on a principled basis an evaluation measure. Chomsky fails to comment on the prospects for this remaining to be the case and fails to spell out the repercussions of such an eventuality for linguistic theory.

Chomsky also has language-specific problems which stand in the way of the principled construction or choice of an evaluation measure:

> The major problem in constructing an evaluation measure for grammars is that of determining which generalizations about a language are significant ones, an evaluation measure must be selected in such a way as to favor these. We have a generalization when a set of rules about distinct items can be replaced by a single rule (or, more generally, partially identical rules) about the whole set, or when it can be shown that a 'natural class' of items undergoes a certain process or set of similar processes. Thus choice of an evaluation measure constitutes a decision as to what are 'similar processes' and 'natural classes'--in short what are significant generalizations. The problem is to devise a procedure that will assign a numerical measure of valuation to a grammar in terms of the degree of linguistically significant generalization that this grammar achieves. The obvious numerical measure to be applied to a grammar is length in terms of number of symbols. But if this is to be a meaningful measure, it is necessary to devise notations and to restrict the form of rules in such a way that significant considerations of complexity and generality are converted into considerations of length, so that real generalizations shorten the grammar and spurious ones do not. Thus it is the notational conventions used in presenting a grammar that define 'significant generalizations,' if the evaluation measure is taken as length [Chomsky 1965 p. 42].

There is a remarkable similarity between Chomsky's 1955 and 1965 statements. The only difference of note between the two passages is that Chomsky has replaced the term "simplicity" with the terms "naturalness" and "linguistically significant generalization" to avoid "confusion" with the philosophy of science sense of simplicity. There is presumably some significant conflict between the predictions that are made on the basis of the philosophy of science notion of simplicity and the predictions made concerning "naturalness" and "degree of linguistically significant generalization" (i.e., in at least one instance the simplest generalization mustn't be the most natural or linguistically significant generalization). Chomsky, however, fails to specify what the cognitive differences are between these three notions and fails to point out a single specific instance where simplicity judgments conflict with judgments of naturalness or degree of linguistically significant generalization. Chomsky's only efforts to define the notions of naturalness and linguistically significant generalization are phrased in terms of the brevity of symbols. This provides no clue whatsoever as to the cognitive content of these notions, since there is no natural relationship between these notions and notational simplicity--only a relationship Chomsky hopes to induce artificially through the judicious selection of notational conventions.

While we may not be clear on what the evaluation measure is supposed to be measuring, we are, nonetheless, constantly being reassured by Chomsky that the choice of an evaluation measure is an empirical determination. Our choice of an evaluation measure can be falsified by showing that it fails to provide a descriptively adequate grammar (see

Chomsky 1965 pp. 25-26). Of course, while in principle this may be possible, in practice it may prove quite difficult since Chomsky has given no independent cognitive characterization of descriptive adequacy. It is tough to come up with a counterexample of a descriptively adequate grammar that is not chosen by an evaluation measure when the only means you have of identifying descriptively adequate grammars is to count symbols using the correct evaluation measure (i.e., the evaluation measure we are attempting to construct).

The last thing I want to point out about Chomsky's discussion of the evaluation measure in Aspects is the role Chomsky assigns the evaluation measure in language acquisition. Chomsky feels that a linguistic theory that satisfies conditions (1)-(5) above provides a language acquisition device:

> This device must search through the set of possible hypotheses G_1, G_2, ..., which are available to it by virtue of condition (iii), and must select grammars that are compatible with the primary linguistic data, represented in terms of (i) and (ii). It is possible to test compatibility by virtue of the fact that the device meets condition (iv). The device would then select one of these potential grammars by the evaluation measure generated by (v) [Chomsky 1965 p. 32].

Chomsky and Halle's The Sound Patternof English (1968) adds a few more refinements and embellishments to their conception of the evaluation measure. The "value" of a rule, sequence of rules, or grammar is now to be the reciprocal of the number of symbols in its minimal representation ("We take the value to be 1/n, where n is the number of symbols, so that the phrase 'higher valued' will have its natural intuitive and numerical meaning" [p. 334]). The notations used for collapsing rules are not counted by the evaluation measure. Chomsky and Halle apparently believe that this is obvious to the point of being self

evident:

> Under any reasonable evaluation measure . . . the notations used in abbreviatory rules do not count in determining the value of the system of rules. . . . These notational transformations are part of the definition of simplicity, and therefore it would be senseless to 'count them' in some way in measuring simplicity [Chomsky 1965 p. 147].

We should note that there is no choice involved in using abbreviatory conventions if we want the evaluation measure to work:

> To say that rules may be given in a simpler form implies that they must be given in that form. More precisely, the notations that we use define a certain valuation for grammars, the value of a grammar is determined by the number of symbols that appear in it when notations are used in the optional fashion [Chomsky 1965 p. 71 fn.].

Chomsky and Halle point out that, while the notations permit us to make statements about a language which are "true and significant" and provide a basis for distinguishing these from other generalizations which are "true and not significant," the notational conventions do not prevent or protect us from making incorrect statements!

In The Sound Pattern of English Chomsky and Halle make the highly significant observation that

> The entire discussion of phonology in this book suffers from a fundamental theoretical inadequacy. Although we do not know how to remedy it fully, we feel that the outlines of a solution can be sketched, at least in part. The problem is that our approach to features, to rules, and to evaluation has been overly formal. Suppose, for example, that we were systematically to interchange features or to replace F by -F (where =+, and F is a feature) throughout our description of English structure. There is nothing in our account of linguistic theory to indicate that the result would be the description of a system that violates certain principles governing human languages. To the extent that this is true, we have failed to formulate the principles of linguistic theory, of universal grammar, in a satisfactory manner. In particular, we have not made any use of the fact that the features have intrinsic content. By taking this intrinsic content into account, we can, so it appears, achieve a deeper and more satisfying solution to some of the problems of lexical redundancy as well as to many other problems that we have skirted in the exposition [Chomsky and Halle 1968 p. 400].

These remarks bring to a close the classic, and more halcyon, period of work on the evaluation measure. They also bring to a close, for the moment, my presentation of Chomsky and Halle's views on the evaluation measure.[11] The theory of markedness (which was developed to address some of the problems Chomsky and Halle allude to in the remarks I just quoted) and other subsequent changes in their thinking will be evaluated as they arise in the context of my critical evaluation of the evaluation measure and the criticism directed against it.

3.2 Motivation for Adopting the Evaluation Measure as a Goal of Linguistic Theory

An obvious place to begin a critical examination of Chomsky and Halle's evaluation measure is with Chomsky's Syntactic Structures discussion of why the goal of providing a practical, mechanical evaluation procedure is a more reasonable--albeit weaker and less desirable--goal for linguists to pursue than the goal of providing a practical, mechanical discovery or decision procedure. There has been a good deal of dissatisfaction in certain quarters with Chomsky's claims about the singular reasonableness and appropriateness of the goal of providing a practical, mechanical evaluation procedure from the moment they were advanced. It is hardly surprising, of course, that scholars who had invested their energies toward constructing a discovery procedure (or some reasonable facsimile thereof) would not accept with equanimity Chomsky's characterization of this goal as unreasonable, unilluminating, and chimerical--particularly since Chomsky could offer no formal proof that this highly desirable goal was in principle impossible to attain. It is somewhat suprising, however, given the depth of dissatisfaction with Chomsky's

conclusions (and the radical changes in focus and methodology they entail), that there have been very few substantive attacks on the logical structure of Chomsky's argumentation in support of the evaluation measure as a goal of linguistic theory.

To recapitulate, Chomsky's argument in Syntactic Structures against the goal of providing a discovery or decision procedure and in favor of the evaluation measure consists of the following series of opinions: (1) he maintains that it is unreasonable to demand that linguistic theory provide anything stronger than a practical mechanical evaluation procedure, (2) he maintains that it is very questionable whether the goal of providing a discovery procedure can be attained in an interesting way (in fact, he feels thatany attempt to provide a discovery procedure will lead to a maze of more and more elaborate and complex analytic procedures that leave unanswered many important questions about the nature of linguistic structures), (3) he maintains that in contradistinction to the goal of providing a discovery procedure, we can, by lowering our sights to the more modest goal of providing a practical evaluation procedure, focus our attention more clearly on the really crucial problems of linguistic theory and arrive at more satisfying answers to them, and (4) he maintains that the goal of providing an evaluation measure, while more modest than that of providing a discovery or decision procedure, is still a significant goal to attain (witness the fact that no other science can boast of having such a procedure in its arsenal). Doubters are assured by Chomsky that the truth of these pronouncements can be verified by the development and comparison of theories propounding the two competing goals (that of providing a discovery procedure and that of providing an evaluation measure).

Since Chomsky's line of argumentation reduces to a series of opinions without supporting evidence, it is not surprising that there have been expressions of outrage in response to Chomsky's proposed change in direction for linguistic theory. One of the few extensive examinations of Chomsky's discussion of the goals of linguistic theory appears in Sanders (1977). Sanders takes issue with Chomsky's discussion on several counts: (1) Sanders (p. 158) notes that Chomsky "presents no evidence in support of this point of view, and no reason at all for supposing that a demand for an evaluation procedure is any less unreasonable than a demand for a discovery procedure or decision procedure," (2) he finds Chomsky's references to "elaborateness," "complexity," "uninterestingness," "weakness," etc., to be beside the point since, for Sanders, the only reasonable standard for a practical procedure is whether or not it gets the job done, (3) he feels that the way to answer questions about the nature of linguistic structure is to construct a theory of natural language, "not a procedure for anything," (4) he finds Chomsky's attempt to establish the evaluation procedure as a goal for linguistic theory by disestablishing the discovery procedure to be a misleading and forceless substitute for argumentation, and (5) he makes the counterclaim that "the fact of the matter is that discovery procedures and evaluation procedures are evidently precisely equivalent both as to relative strength or difficulty of achievement and in their lack of empirical significance" (p. 158).

Sanders is, of course, correct in noting that there is a lack of evidence for Chomsky's claims and his warning that an argument against the validity of "reasonableness" of the discovery procedures as a goal

for linguistic theory is not an argument for the reasonableness and validity of the evaluation measure as a goal is salutary, but it would have been better if he had given us some evidence of the unachievability of the evaluation procedure rather than merely advancing a counterclaim. While the use of metacriteria such as simplicity has yet to be justified to everyone's satisfaction in the philosophy of science, Sanders' claim that the only reasonable standard for a procedure is whether or not it gets the job done is not going to receive wide support. Sanders may not believe that linguistic theory should try to explain language acquisition, but he should admit that it is not sufficient for whoever does have to provide such an explanation to come up with a procedure that does the job (i.e., terminates in a finite amount of time rather than runs on into infinity). The procedure must also "do the job" in "real time." Sanders' rejection of the claim that linguists must provide a model of how children acquire a grammar of their language accounts for his insistence that linguistics need not provide a "procedure for anything." The fact that children invariably succeed in coming up with the highly complex grammars they need to have a command of their language suggests that children are employing an effective procedure of some sort and not relying on luck and imagination.

A somewhat more inventive attack on Chomsky's discussion of the goals of linguistic theory is made by Bruce Derwing in his book, Transformational Grammar as a Theory of Language Acquisition (1973). Derwing argues that Chomsky's attack on the goal of providing a practical, mechanical, discovery procedure is totally vacuous since Chomsky himself sets what is in effect the same goal for linguistic theory by insisting

that linguistic theory provide a language acquisition device (L.A.D.). Derwing asks us to note the fact that the output of both the discovery procedure and the L.A.D. is the one correct grammar of the language.[12] According to Derwing,

> . . . the attempt to specify both an evaluation procedure and the full set of linguistic universals in this way constitutes as much an attempt to provide a 'discovery procedure for grammar' as did any of the earlier more 'mechanical' approaches of the structuralist era. The only differences, clearly, lie in the number (two versus one) and the nature (content versus process) of the components utilized in the attack on the problem, whereas what makes a discovery procedure (in Chomsky's own terms) is not the kind or number of 'prongs' which are used but simply whether or not the ultimate goal is to achieve uniqueness in linguistic description [Derwing 1973 pp. 62-63].

Derwing claims that Chomsky is trying to resolve the content (or Rationalist) vs. process (or Empiricist) debate by falsely claiming that the Rationalist goal is weaker and more reasonable when in fact both have the same goal of providing a method of arriving at the one correct grammar of a language.

Derwing assumes that Chomsky failed to recognize that iterative application of the evaluation measure can, in a finite number of steps,[13] locate the correct (or "simplest") grammar of the language from which the primary linguistic data was drawn just as inevitably and unfailingly as the discovery procedure would.[14] The discovery procedure and the L.A.D. (in which the evaluation procedure is applied iteratively) will both have the correct grammar for the language as their output and, therefore, have and attain the same goal. But Chomsky was aware that the evaluation measure could be applied iteratively to locate the correct grammar and he discusses this possibility in the same chapter of Syntactic Structures that Derwing is criticizing:

> Suppose, for example, that we were to evaluate grammar by measuring some such simple property as length. Then it would be correct to say that we have a practical evaluation procedure for a grammar, since we could count the number of symbols they contain; and it would be literally correct to say that we have a discovery procedure, since we can order all sequences of the finite number of symbols from which grammars are constructed in terms of length, and we can test each of these sequences to see if it is a grammar being sure that after some finite amount of time we shall find the shortest sequence that qualifies. But this is not the type of discovery procedure that is contemplated by those who are attempting to meet the strong requirement discussed above [Chomsky 1957 p. 53].[15]

Unless Chomsky was self-destructing the distinction he was at great pains to make, it would appear from this quotation that the critical distinguishing feature among the procedures is not their output but their manner of operation. Derwing himself admits that it is possible to distinguish cognitively the approaches on the basis of their manner of operation when he diagnoses the evaluation procedure/discovery procedure debate as being a debate between content and process approaches.

Derwing is correct, however, in claiming that Chomsky's discussion of the discovery, decision, and evaluation procedures was misleading. Chomsky tries to create the impression that he is using the terms "strong" and "weak" as they are used in the philosophy of science, where a hypothesis or theory is considered stronger than another if it makes more predictions about the domain it is describing (cf. Goodman's "Safety, Strength, Simplicity"). His diagrams represent a single application of the procedures so that the unwary are led to the conclusion that the discovery procedure is stronger than the decision procedure and decision procedure stronger than the evaluation procedure because we know more after a single application of the discovery procedure than we do after most single applications of the decision procedures, and we know more after a single application of the decision procedure than we know

after a single application of the evaluation measure. A single application of the discovery procedure, by definition, yields the correct grammar, so we can make definite predictions as to what is and what isn't the correct grammar of the language. A single operation of the decision procedure will permit us correctly to partition the set of possible human grammars into the correct grammar for this language and all the incorrect (or not as simple) grammars only if the first grammar it examines happens, by extraordinary coincidence, to be the correct grammar. If not, we can place one grammar in the bucket for incorrect grammars and predict that the correct grammar is one of the remaining grammars. A single application of the evaluation measure will not even permit us to predict whether we have the correct grammar or it remains to be located. We can, of course, throw one of the two grammars compared in the incorrect bucket. But it is not fair to ingore the possibility of constantly reapplying both the decision and evaluation procedures--especially when the L.A.D. is going to rely critically on the iterative application of the evaluation measure. We must look, instead, for some other sense in which the discovery procedure is a stronger procedure than the decision procedure and the decision procedure a stronger procedure than the evaluation measure--some sense of "stronger" related to the manner in which the three procedures operate.

It seems reasonable to call a procedure "stronger" than another if it operates more efficiently (i.e., yields a solution in fewer steps and in a significantly shorter block of time). Such a criterion would be of obvious relevance in the context of language acquisition. While no positive proof can be had at the moment in the absence of a working

discovery procedure, most linguists would be quite willing to grant that a discovery procedure (if one exists) would operate far more efficiently than either a decision or an evaluation procedure since the discovery procedure constructs the grammar directly on the basis of the data while the decision and evaluation procedures must, perforce, wade through a grossly swollen sea of possible human grammars in search of the single optimal grammar. This may, in fact, have been what Chomsky had in mind when he made his "strength" designations, since he commented at several points in his earlier writings on the impracticality of the evaluation measure as a discovery procedure.

There are, however, two potential problems with the above account. First, it's not clear on first inspection that the evaluation procedure, when applied iteratively with a mind to ultimately locating the correct grammar, is an effective procedure. Second, it's not clear that the decision procedure is, in fact, a stronger procedure than the evaluation procedure in terms of efficiency of operation. Let's start with the problem we encounter when we try to apply the evaluation measure iteratively as an effective procedure for locating the correct grammar. The only means Chomsky has ever suggested for guaranteeing that the set of hypotheses (grammars) the child entertains will include the correct grammar of the language is to enumerate the entire set of possible human grammars. We must, therefore, apply the evaluation measure to this set of grammars, one after another, until we locate the correct grammar. If these grammars are enumerated randomly, we will never be sure at any given point in the iterated evaluation process that a more highly valued grammar won't be awaiting us just a bit further on in the

enumeration. We, therefore, cannot stop the iterated evaluation process until we have exhausted the entire set of grammars (something which we cannot do since the set is infinite). What we need, then, is some way of guaranteeing that there will not be a simpler grammar further on in the enumeration. One solution would be to enumerate the set of grammars in descending order of simplicity so that the first grammar that we encounter that is compatible with the data is the correct grammar of the language. This is easier said than done. Since simplicity is a complex notion, it is hard, if not impossible, to find some property that we can use to enumerate the set in descending order of simplicity. But what is true of scientific hypotheses in general is not true of linguistic hypotheses; if Chomsky can devise a set of notations whose brevity faithfully reflects all aspects of simplicity. If this is the case, Chomsky can generate the set of grammars in descending order of simplicity if he can generate them in increasing order of notational complexity. If the grammars are generated in descending order of simplicity, the correct grammar will appear after a finite number of steps if the set of possible human grammars is denumerably infinite.

The problem with demonstrating that the decision procedure is a stronger, more efficient procedure than the evaluation measure emanates from the fact that we are generating the class of possible human grammars in descending order of simplicity. If this is the case both procedures will be equally efficient in that they will both arrive at the correct grammar at exactly the same time since they will simultaneously select the first grammar that turns up in the ordering that is compatible with the data as the correct grammar.[16]

Faced with a situation where the decision and evaluation procedures are obviously equally strong goals, I think that Chomsky would probably decide that the decision procedure is dispensable. I, for one, would welcome this since I have never been able to visualize how the decision procedure is supposed to operate. Just how does the decision procedure recognize the "correct" grammar without benefit of information about how the grammar compares to other grammars in terms of "simplicity" or information about how and on what basis the grammar was constructed? No discussion of the goals of linguistic theory has ever, to my knowledge, presented us with the slightest hint as to what property of a grammar could invariably and unmistakably flag it as the correct grammar.

Let us assume for purposes of argument that we have validated Chomsky's claim that the discovery procedure is "stronger" than the evaluation procedure. We may now ask if it follows from this that the discovery procedure is a "less reasonable" (i.e., more difficult to attain) goal than the evaluation procedure. It is clearly not always the case that the more efficient or otherwise optimal a device is, the more difficult it is to devise. Nor is it the case that the more inefficient, labyrinthian and otherwise nonoptimal a device is, the easier it is to construct. A ranking of devices in terms of optimality (or desirability or efficiency) is in no sense a ranking of these devices in terms of their "reasonableness" (or attainability). This must be determined case by case and will probably vary from individual to individual and on the existence of other knowledge. We cannot say, then, in advance of a careful examination of these two procedures which is likely to be the more reasonable (attainable) goal. In any case, as long as we have not

shown a procedure to be in principle unachievable, we cannot gainsay any scholar who wants to devote his/her time to devising either type of procedure.

In this section, then, we have examined Chomsky's claim that the discovery procedure is stronger than the decision procedure and that the decision procedure is stronger than the evaluation procedure. We have also examined Chomsky's claim that the evaluation procedure is a more reasonable goal for linguistic theory than the decision procedure and the decision procedure a more reasonable goal than the discovery procedure. We have discovered that if we view the decision and evaluation procedures as applying iteratively until they locate the "correct" grammar, these two procedures will be equally efficient. In order for the iterative application of the evaluation procedure to be an effective procedure for locating the "correct" grammar of the language, the set of possible human grammars must be generated in descending order of simplicity or "value." We can do this if we can devise a set of notational conventions that convert considerations of simplicity or value into considerations of notational length and if we can devise an algorithm which enumerates all and only the set of possible human grammars in ascending order of notational complexity (which is equivalent to descending order of simplicity). But if we enumerate our set of hypotheses in descending order of simplicity or value, the first grammar in the enumeration that is compatible with the data will be the correct grammar and this grammar will be immediately (i.e., simultaneously) selected by both the decision and evaluation procedures as the "correct" grammar of the language. Both procedures will thus locate the desired grammar at the same time in all

instances and will be equally efficient at this task. But does this mean that we must now view the two procedures as equally reasonable goals for linguistic theory? It is argued in this section that there is no automatic and invariant correlation between the strength or efficiency of a procedure and its reasonableness or attainability as a goal. We can thus make no judgments about the reasonableness of a procedure as a goal solely on the basis of the strength or efficiency of a procedure. The claim that one procedure will be a more reasonable goal than another must be supported by independent argumentation that refers to the properties of the device rather than to its strength and takes into account the knowledge that is available. Chomsky's claims about the relative reasonableness of the three procedures as goals for linguistic theory were not accompanied by the requisite independent argumentation. We should keep in mind, however, that a goal doesn't have to be demonstrated to be attainable before we work on it. We should also keep in mind that we don't always have to choose the most reasonable of a set of goals. If Chomsky felt that a discovery procedure was attainable (although with somewhat greater difficulty than the evaluation procedure), he would have worked toward that goal. Chomsky has made it clear at several points that he feels the discovery procedure is a more desirable goal.

3.3 The Evaluation Measure: How It Operates and What It Measures

Chomsky's description of the discovery, decision, and evaluation procedures includes the stipulation that they be practical, effective, mechanical procedures. A procedure is mechanical if it consists of a set of completely explicit steps which can be performed without use of creative, intuitive, intelligent outside intervention. A procedure

is effective if it is guaranteed to terminate in a solution (if there actually is a solution) after a finite number of mechanical steps. The requirement of practicality pertains to the time taken to arrive at a solution. The demands of practicality will vary depending on the function served and the technology available. For instance, a procedure may be practical for the linguist but not the child, since the linguist can make use of ultra-high speed computers and the child cannot. These requirements are very stringent and some linguists have asked why we should put ourselves through the paces. As Matthews puts it:

> . . . the real factors are so diverse and numerous (ranging from generality on one side to substantial naturalness on the other) that it is hard to imagine how they might all be captured. . . . Solution a is better than solution b, we are agreed, on the balancing of sundry factors: hence, according to a widespread proposal, we must introduce notations, etc. by which a is 'deemed to be the shorter.' But what purpose does this fiction serve if the real factors can be directly evaluated? [Matthews 1972 p. 391].

While Chomsky has never responded to this particular question directly, it seems reasonable to suppose that Chomsky insisted that these procedures be practical, mechanical and effective for the same reason that he insisted that grammars be practical, mechanical and effective procedures.

Chomsky's main complaint about traditional grammar was that its grammatical descriptions required supplementation by the very linguistic abilities that they were purportedly describing and explaining--in some cases only native speakers could use the grammars without error. If we impose the requirement that grammars consist of formally explicit rules which can be applied in step-by-step mechanical fashion to enumerate all and only the sentences of the language, we can guarantee that we explain rather than utilize the linguistic abilities of native speakers. We will also be clear on what predictions do and do not follow from the

grammatical description. Chomsky points out some of the other virtues of precise formulation in his preface to Syntactic Structures:

> The search for rigorous formulation in linguistics has a much more serious motivation than mere concern for logical niceties or the desire to purify well-established methods of linguistic analysis. Precisely constructed models for linguistic structures can play an important role, both negative and positive, in the process of discovery itself. By pursuing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively a formalized theory may automatically provide solution for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory may have failed to recognize the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by ad hoc adjustments or loose formulations [Chomsky 1957 preface].

It seems reasonable, then, to conclude that Chomsky feels that the evaluation measure should be a formal, practical, mechanical effective procedure because this guarantees that the predictions it makes are explicit and, thus, potentially testable. We will also be able to determine whether or not the evaluation measure is comprehensive (i.e., renders a judgment for all cases). In addition, the existence of a device making such precise and minute discriminations in value will create a pressure to make finer and more accurate judgments and to resolve any inadequacies of formulation.[17]

After hearing this catalogue of the virtues of practical, mechanical, effective procedures, one wonders why no other science is presently working on an evaluation procedure of this sort. The obvious explanation for this is that very severe conditions must be met before such a procedure can be developed. We must have clear judgments of value in all

cases and these judgments must be both empirically supported and notationally mirrorable. And even if we meet these conditions, the evaluation procedure may not totally resolve our indeterminacy problem since there may still be a large number of equally highly valued hypotheses for the data. We will find numerous reasons in this section to believe that linguists have insufficient evidence for the selection of an evaluation measure and to believe that the evaluation measure, if it existed, would not totally resolve our indeterminacy problem.

Chomsky's hopes of providing a mechanical evaluation procedure apparently rested on the observation that simplifications (or generalizations) tend to result in rule and symbol reductions. Chomsky concluded from this that it would be possible to devise a set of notational conventions such that deeper aspects of simplicity, etc., are invariably mirrored in the superficial surface property of notational length.[18] Chomsky felt that it would be acceptable to have some gaps in our judgments of value. Chomsky proposes to cover gaps in our judgments of simplicity or "degree of linguistically significant generalization," etc., by appealing to the clear cases principle. This principle was first enunciated to handle situation where we have no firm judgments of grammaticality. The idea was that when in doubt we would let the grammar decide for us. In other words, whatever the grammar chosen by the evaluation measure predicts about the grammaticality of dubious strings we would accept. Any doubts we have about whether or not we have a natural rule or linguistically significant generalization would be settled by the evaluation measure. In both instances, however, we are assuming that we have found the correct evaluation measure that will choose the correct

grammar for us. Unfortunately, to choose among competing evaluation measures (i.e., evaluation measures selecting different grammars as the correct most highly valued grammar), we require a clear independent conception of what the correct grammar is. Needless to say, since this is the question we need the evaluation measure to solve, we have nothing of the sort. Chomsky and Halle can, of course, always claim that we will choose the simplest of the competing evaluation measures, and this is in effect what they are claiming when they maintain that we will be able to use the clear case principle to cover lacunae in our simplicity judgments. But after two steps into the quicksand there would be few scholars willing to plod further on the assumption that the indeterminacy that afflicts the first two levels will evaporate, leaving us with a single simplest evaluation measure when an intuitive nonmechanical evaluation measure evaluates the purely linguistic evaluation measures.

Once we conclude that if there are unclear cases the linguist will not only be unable to select grammar, but will be unable to select an evaluation measure as well, we naturally want to assess just how badly linguistics is afflicted with unclear cases. Fromkin in her paper, "Simplicity is a Complicated Question," surveys some of the questions we have no clear decisions about:

> Does, for example, a symbol stated in a rule condition cost the same as, double, or half as much as a symbol in a rule? Is the cost of violating the "invariance condition" simply the number of symbols in the rules which convert it to its phonetic representation, or do we count the number of rules which must apply? If the rules are independently justified for other parts of the grammar should the cost be equal to that of say one "simple" rule which applies only to one segment (eg. converts an underlying 'unmarked" /t/ to a dental click)? Is the cost of readjustment rules more or less, or equal to regular P-rules? Do diacritic features cost more than distinctive feature specifications? [Fromkin 1971 p. 83].

To this we may add doubts about whether or not features in the lexicon cost the same as features in the redundancy rules that fill our lexical items (as Chomsky and Halle assume), doubts about Chomsky and Halle's assumption that the more closely related the input and the output of a phonological process are, the greater the value of the rule (in other words, the fewer features changed the more highly valued and natural the rule), doubts as to whether one rule with X features is always equivalent in value to two or more rules containing X features, and doubts about whether or not notational conventions should be costless. Chomsky and Halle freely admit in The Sound Pattern of English (p. 331) that they have no idea what the proper cost of a metathesis rule should be. We must conclude, then, that the linguist is confronted with a significant amount of indeterminacy which adversely affects the linguist's ability to select both a particular grammar for a language and an evaluation measure. This problem can only be solved by finding additional evidence, if it exists, that would give us further insight into what grammar(s) the speakers have internalized.

The fact that the evaluation procedure and its associated notational conventions must be devised on the basis of conclusive empirical evidence has important ramifications for the issue raised by Matthew's query as to why we should bother devising a notation-based mechanical evaluation procedure when we can make all the judgments about generality, naturalness, linguistic significance, etc., directly without appealing to the mechanical evaluation measure. The freedom we have to design notational conventions means that we can devise notational conventions that will shrink grammars to whatever length we deem appropriate (i.e.,

if grammar A is determined to be simpler, etc., than grammar B, we devise conventions permitting us to collapse grammar A to a length shorter than that to which we can collapse grammar B; if they are found to be equally simple, etc., they must collapse to exactly the same length, etc.). The critical step in this process is the empirical determination of which grammar we want our evaluation procedure to select (i.e., which grammar we will devise our notations to make shortest). Without these determinations we would not be able to choose between an evaluation measure with notational conventions that select grammar A and one that selects grammar B (i.e., makes it shortest). The question arises, given the fact that the correlation between notational length and such metacriteria as simplicity and naturalness, etc., is not natural and universal but rather artificially contrived through highly elaborate notational engineering, whether we could ever have confidence in the predictions of the evaluation measure. In other words, could we ever be totally confident that the next grammar won't contain a counterexample requiring us to revise our notational conventions? I don't see how we could ever be sure of this. What this means is that every time we evaluate a new grammar we would have to verify the predictions of our evaluation measure by conducting an independent, empirically based evaluation. If this is the case, why would we also want to conduct an evaluation using our notation-based evaluation measure? Whatever the virtues of formal mechanical procedures may be, they cannot justify the superfluous task of devising a mechanical evaluation procedure to rubber stamp results we must in every instance independently determine.

There is still another problem stemming from the fact that the

correlation between notational length and the judgments of our metacriteria is artifically contrived rather than natural and universal. Let us suppose for sake of argument that the linguist has devised a procedure which will enumerate the set of possible human grammars in ascending order of notational complexity. Chomsky is correct in claiming that there will be only a finite number of grammars of each value (where 'value' is the number of symbols in the grammar--not including the notational conventions) since there are only a finite number of permutations of the finite vocabulary whose symbols add up to the value (e.g., ignoring notational conventions and pretending we could have grammars of value 2 drawn from a nonterminal vocabulary of 2 symbols (a and b), then the maximum number of grammars of value 2 would be 8--four 1-rule grammars with each rule containing 2 symbols (aa; bb; ab; ba) and four 2-rule grammars with each rule containing 1 symbol ([1] a, [2] a; [1] b, [2] b; [1] a, [2] b; [1] b, [2] a). The maximum number of grammars of each value is set by the maximum number of permutations of the finite nonterminal vocabulary that there are adding up to that value. The number of possible grammars of each value increases as the value number goes up (i.e., as the number of symbols goes up and the simplicity, etc., goes down). Chomsky's enumeration will, of course, not include all permutations, since most of these permutations will not be possible human grammars but the number of grammars of each value will increase regularly and dramatically as we ascend in notational complexity.

Let us go back again to the unmarked situation in which notational complexity is not iconic with simplicity, naturalness, etc., and ask ourselves some questions about the set of possible human grammars. How

many simplest grammars are there? How many next to simplest, etc., grammars are there? How do we know that there won't be more grammars of a particular value than Chomsky has possible permutations (if this is the case, then we could not shrink all the grammars of the same intrinsic value to the same notational length)? How do we know that there will be more grammars of value 998 than 997 and more grammars of value 999 than of value 998? I don't see how we can make exact predictions as to how many grammars there are of each value without any empirical investigation whatsoever and I don't see any reason to believe that in no instance will there be more grammars of a particular value than there are permutations for that value. Of course, empirical investigation of the requisite sort may not be possible. Unless we have, in advance, a simplicity (or "value") enumeration of the set of possible human grammars that is not notation based, we could never collect a set of grammars of a particular value and be sure that there wasn't still another grammar of that value further on in the enumeration until we have examined the entire infinite set of possible human grammars--which is obviously not possible. The claim that there will never be more grammars of a particular value than we have permutations is disproven the moment we have assembled more grammars of a particular degree of simplicity than Chomsky has permutations. This can, of course, occur after only a finite number of steps into the enumeration. Suppose Chomsky's enumeration of the set of possible human grammars in ascending order of notational complexity allows for only a thousand grammars of the greatest degree of simplicity. The moment we collect 1001 grammars that we would independently consider to be of the highest degree of simplicity, the claim that we can convert

all considerations of simplicity into considerations of notational length is disproven. Naturally no one is going to undertake such a program of research. No one would have to, since the burden of proof is clearly on Chomsky to provide some evidence that the set of possible human grammars can be partitioned into an infinite number of finite sets such that the first set which contains all the grammars of the highest level of simplicity contains fewer grammars than the second set which contains all the grammars of the second level of simplicity, and the second set contains fewer grammars than the third set which contains all the grammars of the third level of simplicity. It seems implausible that a simplicity partitioning of the set of possible human grammars would result in sets of grammars which become regularly and exponentially larger as we decrease in simplicity if there is no automatic correlation between the simplicity of a theory and its notational brevity.

There are two problems with Chomsky's approach to the evaluation measure that have been incorporated without much comment into the discussion of the evaluation measure in this chapter. The first problem is one that the reader has probably already noted with a certain amount of annoyance. I have been talking about the evaluation measure measuring simplicity, value, naturalness, and degree of linguistically significant generalization without stating what these notions refer to. Actually, there have been no statements in the literature of the cognitive content of these notions. We are told that the simplest, most valuable, most natural, most linguistically significant rule or generalization will (when we perfect our notation-based evaluation measure) be that rule which has the fewest symbols and that the simplest, most valuable

descriptively adequate grammar will be that grammar of the language which contains the fewest symbols. But these definitions tell us how we may someday be able to determine how natural, simple, etc., our rule or grammar is, but they do not tell us what we mean by naturalness, simplicity, degree of linguistically significant generalizations. Without some independent characterization of these notions, how can we possibly determine whether a prediction about the value of some rule or grammar is correct or not?

Let's see, then, if we can determine for ourselves what the evaluation measure must be measuring or selecting. We are operating within a mentalist framework so, assuming that speakers all internalize the same grammar, the evaluation measure which is part of both linguistic theory and the child's language acquisition device must select the grammar that speakers of a language internalize. This grammar will be the most valuable and its rules must be the natural rules and linguistically significant generalizations that Chomsky, etc., are constantly referring to in their writings. The grammar and the rules it contains may not be the simplest grammar and rules for the language since Chomsky, correctly, hasn't wished to claim since the early 1960s that the speaker necessarily comes up with the simplest grammar. It is difficult to advance our understanding of what is meant by a linguistically significant generalizational, natural rule, etc., beyond this point. This is unfortunate since these remarks give the linguist virtually nothing to work with. We have noted that the linguist is cursed with an astronomical number of grammars from a bewildering variety of general theoretical models which correctly predict the linguistic intuitions of the native speaker. The

linguist's task is to decide which grammar the child internalizes and the linguist must do this despite the fact that neither child nor linguist is felt to have valid intuitions about the form of the internalized grammar. There is nothing that we have said about the notions natural rule, linguistically significant generalization, descriptively adequate grammar, etc., that would help the linguist to recognize that one grammar rather than another was the child's grammar and contained the relevant linguistically significant generalizations, natural rules, etc. The evaluation measure and its notational conventions cannot help the linguist since we select as our evaluation measure an evaluation measure which makes the correct predictions about degree of linguistically significant generalizations, etc., and, therefore, to select an evaluation measure we must first have a clear understanding of what are the linguistically significant generalizations, etc., we want it to favor notationally.

If what has just been said is correct, we have to ask why it is the case that Chomsky and many other linguists frequently claim that some rule or another captures a linguistically significant generalization, is a natural rule, etc., without providing any evidence that speakers have internalized this rule. Unless Chomsky, et al., can show some evidence that these rules are the rules that speakers internalize or explain how linguists could have developed valid intuitions of what rules the speakers have internalized, these claims are totally unwarranted.

It is clearly going to be difficult to develop new ways of testing claims about what is and is not a linguistically significant

generalization within the traditional domain of linguistic theory. It is not, however, impossible. Hurford (1977), for instance, feels that statistical analysis can provide us with new means to test claims that have been made about the extent to which some rule makes a linguistically significant generalization. Hurford's strategy is to equate the linguist's term "significant" with the notion "statistically significant." As Hurford points out, linguists are constantly appealing to our intuition that some linguistic state of affairs could not occur "accidentally," "fortuitously," "by coincidence" or "by mere chance." The problem with this is that our intuitions of "likelihood" are often inaccurate and there is no excuse for appealing to these intuitions when statistical procedures are readily available. Hurford's paper demonstrates the use of statistics to determine whether regularities in the data have appeared by mere chance or are significant and in need of explanation. In his paper, Hurford contends that statistical analysis reveals that the correlation of active and passive selectional restrictions is highly significant and, therefore, must be explained while the data traditionally cited as motivating the parenthesis and curly brace notations was found to be nonsignificant.

While the use of statistical procedures seems promising, we should be conscious of the limitations of statistical analysis. The statistical techniques Hurford employs, if valid, identify regularities in the data that are not accidental and need explanation--they do not reveal what the correct explanation for the facts is. How the regularities are to be explained will vary from model to model and within models from linguist to linguist depending on the wealth of solutions the model

makes available. In other words, the use of statistical procedures will narrow down the number of regularities that we have to account for, but they will not narrow down the number of competing explanations (rule systems) for the data that remain to be explained. The indeterminacy problems that have concerned us in this section will, therefore, still plague us. Even with these limitations, Hurford's approach, if valid, would represent an important contribution because we would have a clearer idea of what regularities in the data must be accounted for and what can be ignored.

The second problem with Chomsky's conception of the evaluation measure that has been incorporated into this chapter without comment is the assumption that the evaluation measure selects a single grammar as the correct grammar. Chomsky assumes without argumentation that the child cannot be satisfied with just any grammar in the set of possible human grammars thatmakes the proper predictions, but rather uses the evaluation measure to select a grammar that is in some sense special. Looking at Chomsky's writing from the 1950s, we would probably have assumed that the child was seeking the simplest grammar of the language--possibly the grammar that maximizes ease of learnability, ease of production, and ease of perception.[19] If we gave some thought to the question of whether or not there was for each languageone grammar in the set of possible human grammars which maximized these three factors to a degree that noother grammar did, we would have definitely said no. For one thing, while the goals of learnability, producibility, and perceptibility are sometimes mutually reinforcing and sometimes independent, they are also, in many instances, conflicting. There are, for example,

many cases in historical linguistics where phonetic simplifications resulted in the complication of the phonological system. There is no fixed point of optimal balance between these factors, so there is no single optimal grammar for the language. We would also be unwilling to believe that there is a single simplest grammar if we were talking about objective rather than pragmatic simplicity. The set of possible human grammars will always be too large for there to be a single simplest grammar. Now that Chomsky is no longer claiming that the evaluation measure is selecting the simplest grammar of the language, it is incumbent on him to explain in just what sense the grammar that is selected is special. If the child is not searching for the simplest grammar (in some sense of "simplest"), why is the child searching? Any grammar in the set that makes the appropriate predictions will permit accurate communication with other speakers of the language.

In this section, then, I have tried to determine what the evaluation measure is supposed to be measuring and to determine whether or not we have sufficient empirical evidence to select an evaluation measure. The linguist is confronted with the problem of choosing among a multiplicity of grammars all of which make the correct predictions about the language and the speakers' linguistic intuitions. The evaluation measure is supposed to help us select from this set of competing grammatical descriptions the single descriptively adequate grammar which is internalized by the speakers of the language. The problem with this scenario is that we are faced with indeterminacy in the selection of an evaluation measure as well as a grammar. We cannot choose the correct evaluation measure unless we know what grammars we want our evaluation

procedure to favor. We are thus back to square one trying to figure out which of our set of grammars the child has internalized. If the evaluation measure can tell us nothing we don't already know, the question naturally arises as to why we should bother to construct a notation-based evaluation measure.

Some linguists may be surprised to learn that the linguist is confronted with the problem of selecting a correct evaluation measure. In the 1950s it appeared as if Chomsky was designing the evaluation to measure simplicity in its traditional philosophical sense. If this were the case the linguist's problem would not be the selection of an evaluation measure per se but rather the selection of a set of notational conventions which converted considerations of cognitive simplicity into considerations of notational simplicity (i.e., considerations of symbol length). Moreover we would have at least some legitimate intuitive judgments of the property of grammars our evaluation measure is to be measuring. In cases where we had doubts about the proper judgment of simplicity we could hope for further clarification from the work of philosophers of science on simplicity. It is unlikely that there would be a single simplest grammar for a language, but at least we could narrow down the set of grammatical descriptions of the language that the linguist is trying to choose among (assuming of course we weren't narrowing down a set of grammars of cardinality one to another set of the same cardinality).

Chomsky abandoned the claim that the evaluation measure was measuring simplicity in its "ill understood" philosophy of science sense of the 1960s. We now have no clear conception of what the evaluation measure

is measuring and have a multiplicity of ways of measuring whatever it is that we are measuring. The evaluation measure is supposed to be selecting the correct grammar which is the grammar the child internalizes and which contains the full set of linguistically significant generalizations about the language. Unfortunately, unlike the case where we were measuring simplicity, the linguist has no legitimate intuitions about the grammar he has internalized for his own language or the grammars other speakers have internalized for other languages. In this situation it is hard to see how we can eliminate any of our hypotheses without psycholinguistic evidence which provides us with some insight into the contents of the internalized grammar. It is also hard to see why the child should be searching for some special grammar from the set of predictively accurate grammatical descriptions of the language unless we are told in what sense the grammar is special and why no other grammar would serve the child's purpose as well.

The fact that the evaluation measure is to be notation-based creates a significant problem in and of itself. Whatever it is that the evaluation measure is supposed to be measuring--whether it's simplicity or some other notion of "value," it is unlikely that we can partition the set of possible human grammars into subsets of grammars such that the set of grammars of the highest level of simplicity or value contains fewer grammars than the set of grammars of the second highest value, etc., with each decrease in simplicity or value resulting in a regular exponential leap in the number of grammars.

Yet this seems to be precisely what a notation-based evaluation measure is predicting will be the case. There can be fewer possible

grammars at the first level of simplicity or value than the second and so on because the higher the value of the grammar the fewer the number of symbols and the fewer the possible permutations (i.e., distinct grammars) of that particular value. If in any instance there are more grammars of a certain value than there are distinct grammars of a particular value (i.e., permutations of that number of symbols) this approach is disproven.

3.4 Conclusions

In the mid- and late 1950s, Chomsky advanced the claim that the most reasonable goal for linguistic theory was to provide an evaluation measure. He felt that linguistics, unlike the other sciences, could hope to develop a general, practical, mechanical procedure for choosing among competing grammatical descriptions all of which are compatible with the linguistic data. Careful analysis of the passages in which Chomsky makes this claim reveals no evidence or argumentation that could justify such a claim. Naturally, any linguist is free to pursue whatever research goal seems worthwhile and there is no question that the goal of providing an evaluation measure, if attainable, is a worthwhile goal to pursue--particularly if no stronger goal is attainable. In any case, whatever the merits or attainability of the goal may be, we have been vigorously pursuing it for twenty years and are now in a good position to evaluate just how reasonable the goal of providing an evaluation measure is.

The evaluation measure is supposed to permit us to choose a single special grammar from the astronomical number of grammatical descriptions which accurately predict the linguistic data and the speakers'

intuitions about that data. The grammar that the evaluation measure selects as the correct grammar is supposed to be the grammar that speakers internalize. If we assume that speakers internalize the grammar that is either the objectively or pragmatically simplest grammar of the language, we design a set of notations for our evaluation measure which converts all considerations of simplicity into considerations of notational length which can then be mechanically read and tabulated. We can check the predictions of the evaluation measure against our valid intuitions of simplicity wherever we have them. But Chomsky doesn't wish to claim that speakers internalize the simplest grammar of the language. If speakers are not internalizing the simplest grammar, what properties of grammars are speakers trying to maximize? Chomsky has names for the properties--the rules must be natural and they must make linguistically significant generalizations. The correct grammar presumably has the most natural rules and makes the greatest number of linguisticaly significant generalizations. But what are "natural rules" and "linguistically significant generalizations"? Chomsky doesn't tell us. They seem to be by definition the rules and generalizations made by the speakers of the language in their grammar but we don't know what grammar the speaker has internalized and neither speaker nor linguist has any valid intuitions about what those rules are. Unlike the case where our evaluation measure is measuring simplicity, we have no valid intuitions to help us select an evaluation measure. The only way we can select an evaluation measure is to determine what predictions we want it to make about the correct grammar for languages. We must, thus, have a solution to the problem we want the evaluation measure to solve in order to select the

evaluation measure. The evaluation measure can thus only rubber stamp facts we have already ascertained. If this is the case, why bother devising an evaluation?

The linguist is thus still faced with the problem of deciding which grammar(s) of the set of competing "descriptions" the speakers have internalized. Since the linguistic evidence has all been correctly accounted for by the competing grammars, the linguist must look for evidence from psycholinguistics that would give us some clue as to the properties of the grammar(s) the speakers have internalized. There is no reason to believe that at any time in the near future we will have any clear indication of precisely which grammar(s) the child has internalized so that we could devise an evaluation measure and no reason to devise an evaluation measure once we know this. It is hard to see under these circumstances how the evaluation measure can possibly be considered a reasonable goal for linguistic theory. We will naturally, however, want to continue to search for any evidence as to what grammar(s) speakers internalize. It seems reasonable to suppose, however, that speakers do not all internalize the same grammar. Chomsky has given us no clue as to a property of his target grammar that would justify an elaborate search for this grammar by speakers.

Chomsky has attempted to dismiss references to the indeterminacy problem as mere rediscoveries of the fact that linguistics is an empirical science rather than a branch of logic or mathematics. Since linguistic theories are nontrivial empirical theories, it is said to follow that they are underdetermined by evidence. The problem with this defense is that Chomsky has been asking us to believe that linguistics alone of

the empirical sciences can develop an evaluation measure. Linguistics must therefore be unlike the other sciences with respect to its capacity for internal theory selection. Unfortunately, it turns out that linguistics is indeed like the other empirical sciences in that it has no working procedure for selecting a single theory from a set of competing theories unless it chooses to apply the "ill understood" philosophy of science simplicity criterion or some equally "ill understood" criterion based on pragmatic simplicity. But if we are to apply such criteria we must have some reason to believe that speakers will apply them rigorously and we must be prepared for unclear cases and a multiplicity of equally simplest grammars for a language. We must also find some good reason for trying to devise a set of notations to convert all considerations of simplicity into considerations of notation length.

Footnotes

[1]". . . the most reasonable way to approach the investigation of . . . simplicity . . . seems to be to assume, for some language, that the grammatical sentences are fixed (i.e. that the process of discovery has been completed) and to determine the effect on grammar formulation of explicit considerations of simplicity imposed on the grammatical statement" (Chomsky 1951 p. 31).

[2]This is widely known not to be the case. Philosophers of science are very aware of the fact that two theories A and B which are equally simple conceptually can vary dramatically in their notational complexity depending on the notational system utilized. In one notational system they might be of equal length, in another A may be notationally much shorter while in still another B may be far shorter, and no philosopher of science will bat an eye over this turn of events.

[3]Philosophers of science are concerned not only with simplicity with respect to notational brevity but with economy of theoretical entities (predicates), economy of theoretical premises, and economy of parameters (used in specifying a system).

[4]Chomsky's use of the term "simplicity" in the following discussion also seems to carry its traditional philosophy of science sense:

> "In a sense, then, I think that those linguists who have insisted that a multiplicity of criteria must be brought to bear in each particular aspect of linguistic analysis, are correct, although in another (and more interesting) sense there is only one criterion, over-all simplicity of the grammar, which may in particular instances give rise to a variety of special considerations" (Chomsky 1962b p. 244).

[5]It is also hard to see how Chomsky's comments in his 1975 introduction to The Logical Structure of Linguistic Theory on the influence Goodman's (1951) and Quine's (1953) work on simplicity had on his work are compatible with his claim thathe had not been talking about simplicity in its standard sense:

> ". . . Goodman's ongoing critique of induction seemed to point in a rather different direction, suggesting the inadequacy of inductive approaches. . . . Goodman's investigations of the simplicity of systems . . . also suggested (to me at least) possibilities for a non-taxonomic approach to linguistic theory. Quine's critique of logical empiricism also gave some reason to believe that this line of inquiry might be a plausible one. Quine argued that the principles of scientific theory are confronted with experience as a systematic complex with adjustments possible at various points, governed by such factors as general simplicity.
>
> Perhaps, then, analogous considerations hold for 'the fundamental problem of linguistic theory'" (Chomsky 1955/75 p. 33).

[6]The distinction between discovery and evaluation procedures goes back to The Morphophonemics of Modern Hebrew and the decision

procedure was first mentioned in The Logical Structure of Linguistic Theory.

[7]Cf. Chomsky 1955/75, 1957(b), and 1959 for discussion of specific proposals for discovery procedures.

[8]See Chapter 4 and the relevant sections of the Chapter 1 summary.

[9]Chomsky provides the following alternate phrasing for criteria 1-4: "(i) a technique for representing input signals, (ii) a way of representing structural information about these signals, (iii) some initial delimitation of a class of possible hypotheses about language structure, (iv) a method for determining what each such hypothesis implies with respect to each sentence" (p. 31).

[10]Chomsky provides the following alternate phrasing for criterion 5: "(v) a method for selecting one of the presumably, infinitely many hypotheses that are allowed by (iii) and are compatible with the given primary linguistic data" (p. 31).

[11]I have only presented Chomsky and Halle's thinking about the nature and role of the evaluation measure in synchronic linguistics. Halle, particularly, tried to draw implications for historical linguistics whenever possible. Halle (1961 p. 94) believes that, since the use of rule ordering can provide significant reductions in number and length of rules and rule ordering thus reflects simplicity, it becomes reasonable to propose that the order of the rules also reflects the chronology of their appearance in the language. If this assumption is correct, we have a tool for historical reconstruction that will be particularly useful where no external evidence exists such as written records or borrowing. Halle feels that this principle has already been used implicitly in historical linguistics. He feels that the acceptance of Grimm's and Verner's laws as historical facts is "based wholly on considerations of simplicity" (p. 94). Where simplicity considerations do not dictate an ordering, historical linguistic usually show considerable disagrement as to relative chronology.

The fact that rule ordering can accomplish significant simplifications has still another significant implication for historical linguistics. Halle notes that while the process of constructing the simplest (optimal) grammar capable of generating the set of utterances is clearly part of the child's behavior, this may not be true of adults (who seem almost completely to lose the ability to master new languages like a native). Halle feels that the adults' inability to master new languages like a native is "due to deterioration or loss in the adult's ability to construct optimal (simplest) grammars on the basis of a restricted corpus of examples" (Halle 1962 p. 387). Since the language of the adult and therefore the adult's grammar can and does change, what this means is that "changes in later life are restricted to the addition or elimination of a few rules in the grammar, and that a wholesale restructuring of his grammar is beyond the capabilities of the average adult" (Halle

1962 p. 387). Since the addition of rules with no further restructuring may result in a grammar that is nonoptimal (i.e., not the simplest), it follows that adults and children may have different grammars for the same language.

[12]Chomsky on occasion (usually in footnotes) allows for the possibility that there might be a small set of equally simple grammars.

[13]If simplicity can be successfully defined notationally, if we can enumerate the set of possible human grammars in ascending order of notational complexity, and if there are only a finite number of grammars of any one value.

[14]Albeit, probably more circuitously (i.e., with more steps, greater consumption of time, and more input data [the corpus + the class of possible human grammars vs. just the corpus]) than the discovery procedure.

[15]In fairness to Derwing (and, for that matter, myself since I once made the same assertion as Derwing), the reader should be warned that most of Chomsky's discussion is conducted as if the decision and evaluation procedures were not to be thought of as applying iteratively on the class of possible human grammars. For instance, Chomsky talks of grammars being presented for evaluation after they were constructed in some ad hoc manner on the basis of intuition or other equally unseemly bases. He claims we needn't concern ourselves with where the grammars come from. If Chomsky had been thinking of applying these procedures iteratively to find the correct grammar, he would have recognized that we must be able to guarantee that the evaluation procedure encounters the correct grammar and we must therefore be very concerned with where the grammars the evaluation measure evaluates are coming from. But if this is the case, and Chomsky as I showed had considered the possibility of applying at least the evaluation measure iteratively, then Chomsky deliberately misled the reader into thinking the evaluation measure was a weaker goal. Unless Chomsky had some difference in operation in mind rather than output when he advanced the distinction, the distinction is incoherent.

[16]Note that the evaluation measure would not actually compare grammars in the manner Chomsky diagrams in Syntactic Structures. It is also possible that there would be more than one equally simple grammar. If this possibility is allowed, then the correct grammar(s) would be found through a search of all grammars of the same notational complexity as the first grammar found compatible with the data.

[17]I have done my best to provide an argument for why we should have a mechanical notation-based algorithm and give Chomsky his day in court, but if the truth be known my sentiments are entirely with Matthews. The evaluation measure never tells us anything we do not already know on the basis of empirical evidence. If we don't know in advance what the correct solution is, the evaluation measure cannot tell.

There is no natural relationship between notational simplicity and "simplicity," "linguistic significance" or value as Chomsky is prepared to admit, so the evaluation measure has no predictive force. When in doubt we cannot let the evaluation measure decide for us. With our power to devise notational conventions we are capable of making any of the competing grammars notationally simplest, so it is always a question of having or not having the relevant empirical evidence to help us select our correct notational convention (i.e., our evaluation measure) from a set of competing, conflicting evaluation measures. Whatever the virtues of formalism and mechanical procedures are, they can't possibly outweigh the total pointlessness of trying to rubberstamp something we already know from nonmechanical empirical considerations.

[18]It is important to recognize that Chomsky was not claiming that there was a natural iconic relationship between notational simplicity (i.e., length) and deeper aspects of simplicity (or "value"). He felt that there was a tendency toward such a correlation and hoped or assumed that, within a single model of linguistic theory T.G., he could develop notational conventions such that conceptual and notational simplicity always coincide. Fred Householder was confused about what Chomsky and Halle were claiming when he claimed that Halle believes that "no symbol-system may be judged economical unless it is always the case that the more general statement, when expressed in it, is the shorter" and commented that

> "Where Halle got the idea for this principle, I cannot discover. My friends in various other fields (including philosophy of science) have not heard of it and obvious counter-examples come readily to mind. In arithmetic, for example, a proposition about every odd number is surely more general than one about the number 3, yet the former is written '2n+1 for the integral n' or the like and the latter is '3'" (Householder 1965 p. 16).

[19]Some philosophers may be surprised that factors like these would be included in the evaluation measure since philosophers are not usually particularly concerned with such factors. Hesse (1967 and 1974 Chapter 10) and many other philosophers of science do not find subjective (or pragmatic, or psychological) simplicity (which is said to include considerations like the intelligibility of theories [the ease with which they are understood and imaginatively or mathematically manipulated], ease of computation permitted by a theory, and ease of applicability to existing problems) to be of much philosophical interest. The motivation for Hesse et al. lies in the fact that while there may be a partial correlation between subjective simplicity and types of objective simplicity, the subjective simplicity of theories will vary from person to person and task to task depending on the psychological makeup of individuals and their intelligence or, in the case of machines and mechanical problems, on the logical and material tools available. It will be difficult, then, to make precise statements of any generality about this kind of simplicity and even if an objective definition of convenience, intelligibility, etc., were available, we would have no truly sound basis for

preferring theories which are subjectively simple--since this sort of simplicity is not necessarily correlated with the truth claims of scientific concepts and theories (not that objective simplicity has been conclusively demonstrated to have an unusually strong hold on the truth). A true theory may turn out to be highly inconvenient. But whatever the general philosophical interest of subjective or pragmatic simplicity may be, we must expect to be very involved with such criteria when we attempt to determine what grammars humans internalize.

CHAPTER 4

THE LANGUAGE ACQUISITION DEVICE AS A THEORY OF LANGUAGE ACQUISITION

This chapter reviews and evaluates the standard criticisms of the Language Acquisition Device as a theory of language acquisition before introducing a new line of criticism related to the critical question of whether or not the Language Acquisition Device can locate a target grammar in real time. The chapter will consist of three sections--the first evaluating the existing criticisms, the second introducing a new line of criticism and the third stating conclusions and suggesting directions for further research.

4.1 The Standard Criticisms

Derwing, Matthews, Watt and numerous others have attacked the evaluation measure because they believe that it selects grammars that are (1) not psychologically real, (2) unlearnable, and (3) troublesome to use. They feel that Chomsky's efforts to capture every conceivable regularity and remove every redundancy (that can be removed by a redundancy rule that costs fewer features than it saves) has led to the postulation of increasingly abstract grammars which are connected to surface structure by longer and longer chains of rules and therefore difficult to learn and use. They find no reason to believe that children capture all the regularities or remove all the redundancies that Chomsky and Halle would have us believe they do--particularly since many of the

lexical items and syntactic structures that move us to postulate more and more abstract underlying representations are encountered only infrequently and are learned very late, in fact, learned at a time when Chomsky and Halle indicate the child is quickly losing or has entirely lost the ability to restructure radically his/her grammar.

In evaluating these criticisms, we must recall that they allege defects with a specific formulation of the evaluation measure and do not speak to the question of whether or not we should be seeking to devise an evaluation measure. In other words, even if all these criticisms are borne out, all that would be suggested is that we should be constructing different sorts of grammars and reworking our evaluation measure to select these new grammars (not that this wouldn't be quite a radical change for current theory). The evaluation measure is just a labeling device designed to pick out whatever grammar the speakers of a language internalize (whatever grammar the speakers uniformly internalize is by definition the "correct grammar"). Whether the grammar that speakers internalize is the objectively simplest grammar or the subjectively simplest grammar (i.e., the grammar which maximizes ease of learnability, ease of storage, ease of production and ease of recognition) or even the objectively or subjectively most complicated, unrevealing and byzantine grammar possible for the language, Chomsky maintains that as long as we know what we want selected as the correct grammar we will be able to devise a notation-based evaluation measure that will duplicate our preference.

Needless to say there have been many linguists who have rejected Chomsky's "instantaneous acquisition" idealization. James McCawley

(1967, 1968, and 1974) has been one of the most persistent critics of both the "instantaneous acquisition" idealization and the claim that children succeed in internalizing the optimal grammar. According to McCawley:

> The most serious defect of this account of language acquisition is that it treats the acquisition of language as if it could be divided into two phases, the recording of data and the construction of grammar, which were separate in time. The model thus treats the child as if it were a linguist who goes off to New Guinea, takes down ten notebooks full of data, and doesn't start writing his grammar until he is on the boat back to the United States and has no more informants around. However, it is clear that language acquisition does not work like that; at any stage of the game the child has a grammar of some sort and probably remembers relatively little of the data which he was presented with while he was formulating that grammar. Each stage of language acquisition consists of modifying the grammars which the child has already acquired by making it cover facts which the child has been presented with and his grammar does not yet cover [McCawley 1967 p. 51].

It is at this point that Chomsky and McCawley diverge. According to Chomsky the child confronted with new data incompatible with the internalized grammar would then abandon the grammar and construct the simplest grammar compatible with the data, a grammar that could theoretically differ considerably from the original grammar. According to McCawley, the child will be conservative and try to preserve his/her existing grammar since it has been useful to him/her as a tool of communication. Only those details of it that have proven ineffective tools of communication will be abandoned. The result may or may not be the simplest grammar of the language. McCawley feels this makes particular sense since the child does not have the original corpus of data any longer. Rather than try to recreate that data, why not do a patchup job? He might also add that one of the ways in which grammars are said to be simplified over time is when children resist revising their

grammars to take in additional data complexities that are uncovered late in the course of language acquisition (Kiparsky calls this 'imperfect learning'). If children can resist complicating a rule they can resist undertaking the 'cryptographic' task of total restructuring (if the patchup of the existing grammar doesn't become too cumbersome).

McCawley claims that Chomsky and Halle can vindicate their conception of the evaluation measure if they can demonstrate that

> . . . a measure of distance between grammars could be set up which conformed to the revisions that children actually make in their grammars over the course of time and which had the property that any chain of development that followed the metric would lead to the 'simplest' grammar that conformed to the total set of facts involved in the chain of development, where 'simplest' is definable in terms of the measure of distance between grammars, e.g., the 'complexity' of a grammar equals its distance from the null grammar. However, I see no reason why such a thing as 'distance from the null grammar' should play any role in language acquisition. At any stage in the child's development, his goal is to get a grammar that works and which may be put to work instantly, and there is no reason to suppose that the child would reject a grammar that is 'close' to his present grammar and which 'works,' just because of its distance from the null grammar [McCawley 1974 p. 6].

Chomsky's response to McCawley's attack has been to toss the burden of proof back into McCawley's lap, i.e., to demand that McCawley prove that children do not totally restructure and come up with the optimal grammar. It is hard to see, though, how the burden of proof can be properly shifted onto McCawley's shoulders when Chomsky has yet to cite any communicative disadvantages that would ensue from the child's internalizing a less than perfectly optimal grammar.

Actually, McCawley is giving Chomsky's evaluation measure too much credit when he is accepting that its goal is an optimal grammar. Chomsky claims the evaluation measure is not seeking the "simplest" grammar. The grammar that the evaluation measure is seeking must, then, be

optimal in some other sense. I can't think of what other criterion of optimality Chomsky's evaluation measure can be seeking to maximize and Chomsky hasn't even hinted at what other criterion of optimality he has in mind. If it is difficult to justify the claim that the child is seeking either the objectively or subjectively simplest grammar, it is impossible to justify a search for a grammar whose "specialness" cannot be described in intelligible cognitive terms. Since this debate is not going to be concluded to everyone's satisfaction until we actually know what grammar(s) children internalize, we can move on to a line of criticism that seems much more conclusive.

4.2 Does the Language Acquisition Device Operate in Real Time?

This section seeks to determine whether or not the Language Acquisition Device can locate a target grammar in anything even vaguely approaching real time. If it can't, it cannot possibly be an explanatory model of language acquisition. Fortunately, to examine this question we don't really need an accurate specification of the possible form of human grammars or a working empirically validated evaluation measure or even a clear idea of what sort of grammar the Language Acquisition Device will have as its output--we have only to look at the manner in which the device operates and the number of grammars it must search through. We are therefore free to ignore all the problems we have uncovered in the last three chapters and concentrate our attention on how the Language Acquisition Device operates.

The most obvious way to construe the language acquisition device as an actual working model of how the child goes about coming up with the 'correct' grammar of his/her language is Jerry Katz's account:

> Given this knowledge [of universals] the child can a priori construct the class of possible grammars independently of empirical data. The evaluation procedure enables him to order these grammars on the basis of their complexity. Then making use of the corpus of utterances he is exposed to, he can test the grammars that rank simplest on this ordering by employing the structure assignment algorithm to deduce from the grammars predictions about the phonological, syntactic and semantic properties of the utterances in the corpus. The grammar he eventually adopts is therefore the simplest one compatible with the total sum of his linguistic experiences [Katz 1971 pp. 140-41].

There are two qualifications we have to make with respect to this account. The first qualification is that the set of possible human grammars must be enumerated in descending order of value. If the grammars are enumerated randomly there is no effective procedure that can actually compare all the grammars to place them in their appropriate "value" classes since this would involve comparing an infinite number of grammars. The second qualification concerns the implication that the evaluation measure is selecting the "simplest" grammar compatible with the data. The evaluation measure is selecting the most "valuable" grammar compatible with the data and the only thing we know about the term "value" is that it is not to be confused with "simplicity" in its traditional ill-understood philosophy of science sense. These qualifications out of the way, we can now proceed to determine whether or not the evaluation procedure can operate in real time.

Obviously the critical factor in determining whether or not the Language Acquisition Device will operate in real time is the number of grammars the device will have to consider before arriving at the "correct" grammar for particular languages. We can't state precisely how many grammars will have to be considered in a particular instance, but we can get a very reasonable idea of the magnitude of the search. Let us

consider a highly simplified case where the search starts with the first set of most highly valued grammars and goes through grammars of the first 100 values before the target grammar is located. To simplify the arithmetic we will assume that the simplest grammars can consist of a single symbol, that the nonterminal vocabulary from which the grammars are constructed contains only ten symbols, and that all permutations of symbols are possible grammars. A search of the first 100 values will involve considering $10^1 + 10^2 + \ldots 10^{100}$ grammars. It is obvious from even this grossly oversimplified example that a search involving only a few values is going to be hopelessly impractical. The crucial thing to note about this example and any actual search that the Language Acquisition Device would conduct is that as the search progresses from one value to the next the number of grammars to be considered is increasing expontentially. Mathematicians consider this to be the earmark of an impractical search.

It is difficult to believe that any search might involve wading through $10^1 + 10^2 + \ldots 10^{100}$ grammars, but this clearly must be the case. The set of possible human languages is said to be infinite. Let us consider the most favorable enumeration possible--one in which the enumeration enumerates the "optimal" grammar for each of these languages uninterrupted by the appearance of a second grammar for any of these languages. There still must be some language whose grammar will not be enumerated until $10^1 + 10^2 + \ldots 10^{100}$ other grammars have been enumerated. It may even be the case that English is the grammar that will be enumerated after $10^1 + 10^2 + \ldots 10^{100}$ other grammars have been enumerated.[1]

The question that naturally comes to mind once one has thought about the operation of the Language Acquisition Device in some detail is why didn't Chomsky realize how hopelessly impractical this procedure is. After rereading Chomsky's discussions of the evaluation measure, one discovers that Chomsky in fact was aware as early as 1955 that the use of the evaluation measure as a discovery procedure is impractical:

> Suppose that we have constructed linguistic theory in such a way that given a corpus of utterances for which we know in advance that there is some grammar, it is the case that (i) all systems having the prescribed form of grammars given the corpus, can be enumerated in order of increasing complexity; (ii) given any such system, it is possible to determine in a mechanical way whether the generated set of derivations provides a proper level structure. In this case, the general theory will provide a discovery procedure for grammars, though of course not a practical one [Chomsky 1955/75 p. 116].

But if Chomsky is aware that the procedure is impractical, why is he claiming that it is an explanatory model of language acquisition? Chomsky must have something else in mind that at least in principle speeds up the procedure. We get a clue to what else Chomsky has in mind in his 1962 paper, "Explanatory Model in Linguistics," where he indicates that

> To make this model realistic (that is, to make the language-learning device practical) we must supply it with some sort of heuristic or inductive principles that enable it, given, input data, to make a rapid selection of several potential grammars to be submitted to the procedure of evaluation. One might include in (8) a specification of such heuristic procedures, or one might argue that they are not a reasonable concern of linguistic theory. In any event, it seems to me that the scope and effectiveness of heuristic, inductive procedures has been greatly exaggerated. To select a recursive, formalized grammar, given fragmentary data, a learning device must obviously contain both heuristic procedures and a specification of the form of grammars as part of its internal structure. But the task remaining to heuristic procedures is obviously lightened as we make the specification of the form of grammars increasingly narrow and restrictive. It seems to me that the relative suddenness, uniformity, and universality of language learning, the bewildering complexity of the resulting skills, and the subtlety and finesse with which

> they are exercised, all point to the conclusion that a primary and essential factor is the contribution of an organism with highly intricate and specific initial structure [Chomsky 1962 pp. 535-36].

Chomsky's efforts to make the Language Acquisition Device a practical procedure for locating the correct grammar are, to say the least, highly problematical. We have a model of language acquisition which by any stretch of the imagination is hopelessly impractical, yet Chomsky wishes to advance it as a model of language acquisition. All Chomsky can tell us to make us believe that this model can be made practical is that there is some set of heuristics that will make the model practical. He can't tell us what they are and even hints that it might not even be the province of the linguist to provide such a specification. What reason has Chomsky provided to make us believe that his Language Acquisition Device can be an explanatory model of language acquisition? I can see none. What we seem to have is a recipe for transmuting any false theory into a plausible one. All the scientist has to do to save any theory is to claim that the theory can be supplemented by some totally unspecified set of heuristics (or hypotheses) that will make the theory or model realistic. The critics of the theory are supposed to accept on faith that such heuristics exist since the scientist is under no obligation to ever come up with even a single helping heuristic because this is the task of some other discipline. We are also supposed to believe that the other discipline which provides the heuristics that are to turn a hopelessly inaccurate model into an explanatory model is performing only a small service for the linguist since it seems to Chomsky that "the scope and effectiveness of heuristic, inductive procedures has been greatly exaggerated."

I don't see that it is unreasonable to insist that Chomsky provide us with the set of heuristics that will turn a hopelessly impractical model into a realistic explanatory model of language acquisition before we believe that this can be done and before we believe that it tells us something about the mind and its operation. I don't see that it is unreasonable to insist that we call a set of inductive techniques that invariably provide the speaker with the means to develop a grammar of the language an algorithm rather than a set of heuristics (which implies that in some instances the inductive procedures will fail). I don't see that it is unreasonable to claim that the importance of such an algorithm could hardly be exaggerated. I do find it unreasonable, however, to claim that such an algorithm exists.

To see why it is unreasonable to claim that such an algorithm exists we need only return to our consideration of the most optimal enumeration possible of the set of possible human grammars. This enumeration is not interrupted by the appearance of a second grammar for any language until the optimal grammar for every language is enumerated (which would be never). What clue could we have as to which step in the enumeration will produce the grammar for the language we are interested in? And if we did know that the grammar would be the 100 billionth to be enumerated, how could we make our enumeration algorithm crank that grammar out first? To see why it is impossible to claim that such an algorithm exists, we need only recall from Chapter 1 that E. Gold demonstrated in 1967 that no algorithm can uniformly locate a target faster than the notoriously inefficient enumeration algorithm.

4.3 Conclusions

The obvious conclusion of this chapter is that Chomsky's Language Acquisition Device is not an explanatory model of language acquisition. Since Chomsky's Language Acquisition Device cannot possibly be an explanatory model of language acquisition, we will have to search for a different model which does not incorporate the debilitating features of Chomsky's model. The problem with Chomsky's model is that it uses an enumeration procedure to locate the correct grammar. No model that uses an enumeration as a source of its hypotheses can function in real time. What we need, then, is an algorithm which contains a set of inductive procedures which enable us to construct the grammar directly on the basis of the linguistic data. We will probably want to call this algorithm a discovery procedure, but we should keep in mind that we are not looking for just a set of techniques for the segmentation and classification of linguistic elements. While we are on the topic of what we are not looking for, I feel it is unreasonable to insist that our inductive procedures pick out some single grammar. Any grammar that is a possible human grammar and makes the accurate predictions needed for communication is fair game for the discovery procedure. There will be a tendency to develop the most pragmatically simple grammar that is convenient, but there will be no imperative to find some single special grammar. One of the problems with the evaluation measure was that it made it difficult to claim that any grammars other than the target grammar were possible human grammars. If we enumerated the set of possible human grammars in descending order of "value," it was in principle impossible for any grammar other than the optimal grammar to be acquired (if any grammar at

all could be acquired using this method). If these other grammars can't be acquired, how can they be possible human grammars in any meaningful sense?

Since our "discovery" procedure must enable us to acquire any possible human grammars, we still need to have a precise designation of the set of possible human grammars so that we can make sure that our "discovery" procedure can acquire all and only these grammars. The information we were seeking for our enumeration algorithm we will need for our discovery procedure. None of the research that has been conducted on the possible form of human grammars need go to waste. The principles involved in the algorithm may or may not be learnable. This remains to be seen. Until we see we must reserve judgment as to whether or not this algorithm is innately specified.

Footnotes

[1]When we evaluate the practicality of Chomsky's Language Acquisition Device it is important to recall that the child is viewed as using the L.A.D. not once but every time additional data calls for a change in the grammar. I agree with McCawley that it is unreasonable to suppose that a child will abandon rather than patch up an existing grammar when confronted with data that calls for a revision in the form of the grammar. Chomsky insists, however, that the child will totally restructure and come up with the "correct grammar for both the new and old data." Since the child encounters new data that is not adequately handled by his existing grammar on numerous occaions, our estimate of the impracticality of Chomsky's model must be suitably increased.

CHAPTER 5

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

In the late 1950s and early 1960s Chomsky proposed that linguistics become a mentalistic science whose goal it is to describe the linguistic competence of native speakers and explain how this competence is acquired. Chomsky believed that it was both revealing and accurate to view the linguist who is constructing a grammar and the child who is acquiring a grammar as performing analogous tasks. Chomsky was convinced that the child brings to the task of language acquisition an innately specified knowledge of the form of grammar which is the equivalent of the linguist's general linguistic theory with its specification of the organizational makeup of grammars (i.e., their components and their interaction), its specification of the formal properties of the rules contained within these components, and its specification of the theoretical vocabulary utilized in these rules (e.g., phonetic and semantic features, grammatical categories, etc.). The linguist has a corpus of data and the child the primary linguistic data he is constantly exposed to during language acquisition. Finally, the linguist and (by hypothesis) the child have an evaluation procedure for selecting a single correct grammar from among the set of grammars compatible with the data. If the linguist's set of organizational, formal and substantive universals and his evaluation measure are the same as the child's, the linguist and the child will construct the same grammar for the same set of data

and the linguist will know what grammar speakers have internalized and how that grammar has been acquired. Chomsky argued that the grammars children acquire are transformational grammars. Chomsky's claims have generated intense interest in psychology and inspired a great deal of research over the last two decades.

In Chapter 1 of this study we examined the question of whether or not Chomsky was correct in claiming that linguistics should be a mentalistic science. In recent years there have been a number of anti-mentalist arguments advanced. All of these arguments against a mentalistic linguistics were evaluated and found wanting. Linguists can attempt to investigate both language and knowledge of language without any conceptual confusion ensuing. It is impossible to conceive of a researcher successfully studying knowledge of language without having a deep and sophisticated understanding of the structure of language. The linguist, then, is in as good a position as the psychologist to study speakers' knowledge of their language. There is no reason for the linguist to claim thatpsycholinguistic or neurolinguistic tests can disprove grammatical rules. These tests falsify claims about what rules are contained in the grammars that speakers have internalized, not claims about what is a possible rule description for some grammatical facts. The mentalist will freely admit that a rule that is not psychologically real may make accurate predictions about grammatical data and be part of a grammar which makes accurate predictions about the language and speakers' linguistic intuitions. No evidence has been provided that suggests that the grammars speakers internalize are in any meaningful sense less optimal or revealing descriptions of the language than other

grammatical descriptions which speakers cannot internalize. There is no good reason, then, why linguistics should not be a mentalistic science and many reasons to believe that the research goals of mentalist linguists are highly significant and of interest to researchers in many disciplines.

The second topic of concern in Chapter 1 was the question of whether or not the grammars children internalize are transformational grammars. The psycholinguistic tests that have been conducted reveal that no version of transformational grammar that has been tested plays an active role in language use. Analysis of alternative roles for the transformational grammar that would permit us to say that the transformational grammar is psychologically real suggests that there is no other viable role that the transformational grammar could be playing. No grammar of any type which is functioning as a random generating device can play a role in a model of the speaker-hearer's linguistic competence since there is no way of accessing data from this device that is consistently faster than the hopelessly impractical enumeration technique. The reasonable conclusion to draw from these studies is that Chomsky has seriously underestimated the significance of his "performance" model and its "heuristics." The set of heuristics whose study Chomsky has continually deferred until a date when we have a working "competence" model is actually a full-fledged algorithm which is as abstract and idealized as the transformational grammars Chomsky has been advocating as a "competence" model. It seems to me that the first order of business for linguists will be to develop a working performance model. There is no evidence whatsoever that the "performance" algorithm will need to be

supplemented by another grammar. As we noted earlier, the analysis of the other roles that have been suggested for transformational grammars reveals that none of these roles is viable. Learnability studies by automata theorists suggest that the algorithms we postulate for our performance models must be no greater in power than primitive recursive automata. We can therefore ignore all general linguistic theories whose grammars have the power of unrestricted rewriting systems when we are searching for a "performance" model.

Chapters 2, 3, and 4 evaluated the merits of Chomsky's Language Acquisition Device as a model of language acquisition. In Chapter 2 we examined what progress has been made toward providing an accurate description of the characteristics of a possible human grammar and examined the role this universal specification is supposed to play in explaining language acquisition. Analysis reveals that our selection of a "universal grammar" is grossly underdetermined by the standard linguistic facts that must be handled by the grammars provided by our "universal grammar" (or general linguistic theory). There are an astronomical number of general linguistic theories whose grammars are capable of making the proper predictions about the linguistic data and speakers' linguistic intuitions. Any claim that we have settled on a single correct characterization of the possible form of human grammars is hopelessly premature. This situation is not, however, cause for surprise or excessive alarm since no other empirical science can boast of a much better record in so short a timespan. To reduce the indeterminacy problem to more manageable proportions, we must reduce the power and abstractness of our linguistic theories and seek additional types of

evidence of relevance to the question of what grammars speakers can internalize. We obviously can't afford to ignore evidence from psycholinguistics. In addition to the problem of deciding on a correct specification of the set of possible human grammars, there is some question as to whether this universal specification plays the critical role Chomsky ascribes to it in explaining how language is acquired. Chomsky repeatedly claims that the more constraints we place on the form of grammars, the fewer hypotheses there are for the child to consider and therefore the easier it is to acquire a language. The problem is that the set of possible human grammars is a denumerably infinite set both before and after each of these constraints is added. It is hard to see how these constraints make the task of language acquisition significantly easier. In any case, even if each additional constraint did make the task of language acquisition easier, we can't just tack on constraints on the sole ground that they are easing the task of language acquisition. We can only add a constraint if it accurately characterizes a limitation on the set of possible human grammars.

Chapter 3 examines the evaluation measure. Chomsky claims that linguistics, unlike the other sciences, can reasonably hope to develop a practical, effective, mechanical notation-based evaluation procedure for choosing among competing linguistic theories at the grammar level and that this is the most reasonable goal for linguistic theory. Careful analysis of Chomsky's argumentation reveals no reason to believe that the evaluation measure is the most reasonable goal for linguistic theory or even a reason to believe that it is a reasonable goal. Analysis of the criteria that must be met before a notation based evaluation measure

can be constructed bears out the suspicion that the goal of providing a notation-based evaluation measure is not a reasonable goal for linguistic theory. Chomsky provides us with no insight into what property our evaluation measure is measuring other than "correctness." The correct grammar seems, by definition, to be whatever grammar the speakers internalize so we can have no idea what we mean by "correctness," "naturalness," etc., unless we know what grammar the speakers internalize. We don't know this and the whole point of the evaluation measure is to spare us the burden of having to know this. Unless the speakers are selecting a grammar on some principled basis and we can spell out what that principled basis is, we can't devise an evaluation measure without knowing in all instances what grammar we want the evaluation measure to select. There are some linguists who make pronouncements about whether or not some rule is natural or makes a linguistically significant generalization, but there is no independent cognitive characterization of the notion "natural rule" and "linguistically significant generalization" that could form the basis for such judgments and these linguists must explain how they could have developed valid intuitions about the form of internalized grammars. In any case, even these linguists will admit that there are numerous areas in which they have no intuitions whatsoever as to what solution the evaluation measure should be selecting. Since we don't know what solution should be selected, we can't choose an evaluation measure from among the set of evaluation measures that are making conflicting predictions about what grammar should be selected. It is not reasonable to suppose that the unclear cases are going to become suddenly less umbrous in the foreseeable future. It is hard to see,

then, how we can reasonably hope to devise an effective, mechanical evaluation procedure.

It is also not even clear that we want a notation-based evaluation measure since notation-based evaluation measures seem to make some very counterintuitive claims about the set of possible human grammars. A notation-based evaluation measure predicts that the number of grammars will increase at a regular exponential rate as we step down each notch in value. Where is the evidence for such an assumption? The number of grammars of any one notational value is finite. If our notion of "value" corresponded to "simplicity" would we be willing to claim that there would not be an infinite number of grammars at any one level of simplicity or that there would never be a case where the number of grammars at some level of simplicity exceeded the number of possible distinct grammars of the designated number of symbols? Whatever Chomsky is measuring in connection with his notion "value," is it likely that there won't be an infinite number of grammars of any one value or at least more grammars than there are distinct permutations of symbols? It seems reasonable to conclude on the basis of the analysis in Chapter 3 that linguists should not be wasting their time trying to devise a practical, mechanical notation-based evaluation measure. It is just not a reasonable goal for linguistic theory. An alternative goal will be suggested as a result of the conclusions of Chapter 4.

Chapter 4 ignores the problems that the earlier chapters diagnosed in connection with specifying the correct set of possible human grammars and the correct evaluation measure and focuses on just the question of whether or not a language acquisition device that has an

enumeration of the set of possible human grammars in descending order of value and an evaluation measure can possibly explain language acquisition. The conclusion of this chapter is that a model of this sort cannot explain language acquisition because the enumeration is hopelessly impractical as a method of locating the most highly valued grammar of a language and there are no heuristics that can speed up the process. A model of this sort cannot, in principle, meet the "real time" restrictions imposed by language acquisition. If one were to specify what sort of heuristics would be needed, it would turn out that the heuristics whose significance Chomsky continually downplays would have to constitute a discovery procedure (or algorithm).

The conclusion of Chapter 4 is that there must be a major switch in our search for an explanatorily adequate model of language acquisition. We must focus our research on developing an algorithm which constructs a grammar directly on the basis of the data rather than searches for a grammar that is embedded in an enumeration of the set of possible human grammars. It seems unreasonable to insist that this algorithm construct a single correct grammar of the language. Any grammar that is a possible human grammar and accurately predicts the data and speakers' linguistic intuitions will be communicatively adequate and fair game for the algorithm. There may be a tendency to select the most pragmatically simple grammar convenient but this will be a tendency, not an imperative. One would not expect a child to restructure totally in order to get the "simplest" grammar when some patchwork on the existing grammar will do fine. There's also no reason to expect that all speakers will arrive at the same grammar. If they did it would be hard to see in

what sense the other grammars that also accurately predict the linguistic data and speakers' linguistic intuitions would ever in principle be possible human grammars for the language in question. Since our algorithm must construct all and only the possible human grammars for any set of data, we will still need to know what is and is not a possible human grammar. This means that our research on linguistic universals will not go to waste but it also means that an accurate discovery algorithm will not be forthcoming in the near future.

There are a few things that we can state about the discovery algorithm. First, and most obvious, is the fact that its inductive technique will not be the standard structuralist techniques of data segmentation and classification. Mathematicians and psychologists have been working on different sorts of inductive techniques for the last two decades and we will have to evaluate what progress has been made since linguists last concerned themselves with inductive procedures. The second thing we can state on the basis of the learnability studies of automata theorists is that no constructive technique will permit us to learn grammars that are more powerful than primitive recursive automata. The last thing we can state about the discovery algorithm is that until we develop it we cannot ascertain whether its principles are learnable or must be innately specified. My own guess is that Chomsky is correct in believing that there is an innately specified language faculty. There does seem to be independent evidence that child and adult do not approach language learning in the same way. It would be very odd if the adult's techniques for learning a second language are inferior to the child's if these are learned.

The conclusions of this study should be placed in perspective. If correct, these conclusions suggest that Chomsky was wrong about the precise nature of the grammars that children internalize and the device they use to acquire the grammars they do internalize. Being wrong seems to be the inevitable fate of one who takes the risk of advancing a theory of any scope. We should be thankful that Chomsky had the intellect and the guts to suggest important new directions of research at almost every level of linguistic theory. Chomsky's positive contributions to the field are immense and dwarf anything that most of his critics can ever hope to accomplish with their own work. I don't feel that we linguists have very much of a definitive nature to state to psychologists, philosophers, etc., but there is no doubt that Chomsky's work and insights have advanced us a quantum leap in the direction of being able to make a significant contribution to other disciplines and to our understanding of the human mind and its cognitive capacities.

BIBLIOGRAPHY

Bever, T. 1968. A survey of some recent work in psycholinguistics. In W. Plath (ed.) 1968, pp. 1-66.

Carr, J. W. (ed.). 1959. Computer Programming and Artificial Intelligence. Michigan: University of Michigan Press.

Chomsky, A. N. 1951. Morphophonemics of Modern Hebrew. Unpublished Master's thesis, University of Pennsylvania.

_____. 1955/75. The Logical Structure of Linguistic Theory. New York: Plenum.

_____. 1957. Syntactic Structures. The Hague: Mouton.

_____. 1958. Linguistics, logic, psychology and computers. In J. W. Carr (ed.) 1959, pp. 429-54.

_____. 1959. Review of Verbal Behavior, by B. F. Skinner. Language, 35, 26-58.

_____. 1961 On the notion rule of grammar. In R. Jakobson (ed.) 1961, pp. 6-24.

_____. 1962a. Explanatory models in linguistics. In E. Nagel, P. Suppes, and A. Tarski (eds.) 1962, pp. 528-50.

_____. 1962b. A transformational approach to syntax. In J. Fodor and J. Katz (eds.) 1964, pp. 211-45.

_____. 1964a. Current Issues in Linguistic Theory. The Hague: Mouton.

_____. 1964b. The logical basis of linguistic theory. In H. Lunt (ed.) 1964, pp. 978-1008.

_____. 1964c. Formal discussion. In M. Lester (ed.) 1970, pp. 41-50.

_____. 1965. Aspects of the Theory of Syntax. Cambridge, Mass.: MIT Press.

_____. 1966. Linguistic theory. In M. Lester (ed.) 1970, pp. 51-60.

_____. 1967. The formal nature of language. In A. N. Chomsky 1972a, pp. 115-60.

_____. 1970. Deep structure, surface structure, and semantic interpretation. In A. N. Chomsky 1972b, pp. 62-119.

_____. 1972a. Language and Mind. Enlarged Edition. New York: Harcourt Brace Jovanovich.

_____. 1972b. Studies on Semantics in Generative Grammar. The Hague: Mouton.

_____. 1974. Interview with Herman Parret. In H. Parret 1974, pp. 27-54.

_____. 1975. Reflections on Language. New York: Pantheon.

_____. 1976a. On the biological basis of language capacities. In R. W. Rieber (ed.) 1976, pp. 1-24.

_____. 1976b. On the nature of language. In S. Harnad et al. (eds.) 1976, pp. 46-60.

_____. 1977. Filters and control. Linguistic Inquiry, 8, 425-504.

Chomsky, A. N., and G. Miller. 1963. Introduction to the formal analysis of natural languages. In R. D. Luce et al. (eds.) 1963, pp. 269-321

Derwing, B. L. 1973. Transformational Grammar as a Theory of Language Acquisition. London: Cambridge University Press.

Edwards, P. (ed.). 1967. The Encyclopedia of Philosophy. New York: Macmillan Publishing Co.

Fodor, J., T. Bever and W. Garrett. 1974. The Psychology of Language. New York: McGraw Hill.

Fodor, J., and W. Garrett. 1966. Some reflections on competence and performance. In J. Lyons and R. J. Wales (eds.) 1966, pp. 135-79.

Fodor, J., and J. Katz (eds.). 1964. The Structure of Language. Englewood Cliffs, N.J.: Prentice-Hall.

Frege, G. 1967. The Basic Laws of Arithmetic. Berkeley: University of California Press.

Fromkin, V. 1971. Simplicity is a complicated question. U.C.L.A. Working Papers in Phonetics, 19, 61-68.

Gold, E. M. 1967. Language identification in the limit. Information and Control, 10, 447-74.

Goodman, N. 1951. The Structure of Appearance. Cambridge, Mass.: Harvard University Press.

_____. 1961. Safety, strength, simplicity. Philosophy of Science, 28, 150-51.

Halle, M. 1961. On the role of simplicity in linguistic descriptions. In R. Jakobson (ed.) 1961, pp. 89-94.

_____. 1962. Phonology in a generative grammar. In V. Makkai (ed.) 1972, pp. 380-92.

Harnad, S., et al. (eds.). 1976. Origins and Evolution of Language and Speech. New York: The New York Academy of Sciences.

Hayes, J. R. (ed.). 1970. Cognition and the Development of Language. New York: Wiley.

Hesse, Mary. 1967. Simplicity. In P. Edwards (ed.) 1967, vol. 7, pp. 445-48.

_____. 1974. The Structure of Scientific Inference. Berkeley: University of California Press.

Householder, F. W. 1965. On some recent claims in phonological theory. Journal of Linguistics, 1, 13-34.

Hurford, J. R. 1977. The significance of linguistic generalization. Language, 53, 574-620.

Jakobson, R. (ed.). 1961. Structure of Language and Its Mathematical Aspects. Providence, R.I.: American Mathematical Society.

Katz, J. 1971. The Underlying Reality of Language. New York: Harper Torch Books.

Lester, M. 1970. Readings in Applied Transformational Grammar. New York: Holt, Rinehart and Winston.

Luce, R. D., R. Bush and E. Galanter (eds.). 1963. Handbook of Mathematical Psychology: Vol. II. New York: Wiley.

Luchins, A. S. and E. H. Luchins. 1965. Logical Foundations of Mathematics for Behavioral Scientists. New York: Holt, Rinehart and Winston.

Lunt, H. (ed.). 1964. Proceedings of the Ninth International Congress of Linguists, Cambridge, Mass., 1962. The Hague: Mouton.

Lyons, J., and R. J. Wales (eds.). 1966. Psycholinguistics Papers. Edinburgh: Edinburgh University Press.

Makkai, V. B. (ed.). 1972. Phonological Theory. New York: Holt, Rinehart and Winston.

Matthews, P. H. 1972. Inflectional Morphology. London: Cambridge University Press.

McCawley, J. D. 1967. Can you count pluses and minuses before you can count? Unpublished Ms.

_____. 1968. Review of Current Trends in Linguistics, ed. by T. A. Sebeok. Language, 44, 556-93.

_____. 1974. Acquisition models as models of acquisition. Unpublished Ms.

Miller, G. A. 1965. Some preliminaries to psychologists. American Psychologist, 20, 15-20.

Miller, G. A., and A. N. Chomsky. 1963. Finitary models of language users. In R. D. Luce et al. (eds.) 1963, pp. 419-92.

Minsky, M. 1967. Computation. Englewood Cliffs, N.J.: Prentice Hall.

Nagel, E., P. Suppes and A. Tarski (eds.). 1962. Logic, Methodology, and Philosophy of Science. Stanford: Stanford University Press.

Parrett, H. 1974. Discussing Language. The Hague: Mouton.

Peters, P. S. 1970. Why are there so many "universal" bases? Papers in Linguistics, 2, 27-43.

Peters, P. S., and R. W. Ritchie. 1969. A note on the universal base hypothesis. Journal of Linguistics, 5, 150-52.

Plath, W. J. (ed.). 1968. Specification and Utilization of a Transformational Grammar. Yorktown Heights, N.Y.: IBM Corporation.

Quine, W. V. O. 1953. From a Logical Point of View. Cambridge, Mass.: Harvard University Press.

Rieber, R. W. (ed.). 1976. The Neuropsychology of Language. New York: Plenum.

Sanders, G. A. 1977. Some preliminary remarks on simplicity and evaluation procedures in linguistics. Minnesota Working Papers in Linguistics and Philosophy of Language, 4, 155-67.

Steinberg, D. D. 1970. Psychological aspects of Chomsky's competence-performance distinction. University of Hawaii Department of Linguistics Working Papers in Linguistics, vol. 2 (2), 180-92.

Watt, W. C. On two hypotheses concerning psycholinguistics. In J. R. Hayes (ed.) 1970, pp. 137-220.