Mathematical Transformations in Syntax

A thesis submitted in partial fulfillment of the requirements
For the degree of Master of Arts in Linguistics

By
Rebecca Anne Santini

December 2015
The thesis of Rebecca Santini is approved:

______________________________________   ___________
Dr. Chadwick Sprouse       Date

______________________________________   ___________
Dr. Enchao Shi        Date

______________________________________   ___________
Dr. Sharon M. Klein, Chair       Date

California State University, Northridge
ii
# Table of Contents

Signature Page ii
Abstract iv

Chapter 1: Introduction 1
Chapter 2: Generative Grammar: Some Foundations 15
Chapter 3: Generative Grammar Revised 35
Chapter 4: Shifts 46
Chapter 5: Consequences 56
Chapter 6: Principles and Parameters 70
Chapter 7: The Minimalist Program 79
Conclusion 97
References 103
Appendix: Definitions 107
This thesis explores the application and consequences of mathematical transformations in the history of Syntax. The motivation for including mathematics in linguistics is discussed along with a brief survey of the state of linguistic theory before generative grammar. Generative grammar and the introduction of mathematical transformations is introduced and the application of mathematical transformations is traced from that point as the model developed continuing with generative semantics, Principles and Parameters and finally the minimalist program.
Chapter 1: Introduction

1.0 Introductory Remarks

The nature of the human language faculty has, since the first offerings of generative grammar, been conceptualized as a computational system with fairly restrictive characteristics, but with an infinite capacity for combining structures associated with meanings. Mathematical modeling has been at the core of generative grammar and its theoretical framework. In theoretical mathematics, numbers are not studied. Instead, the focus is purely structural the number system being an effective way of exemplifying the structures being studied. The structures are what is important, and exist with or without the number system. Likewise, syntax\(^1\) is essentially the examination of the theoretical structure of human language. Syntax seeks to develop a model of structure consistent with observable speaker judgments, but just like math without numbers, the structural system developed by syntax exists without real world usage. In order to explore syntax from this perspective, we will look at the definitions of sets and transformations in mathematics and we will examine the source of their application to linguistics throughout the progressive development of the generative paradigm. Along the way, we will also explore the motivations that led to each of the shifts in the development of the paradigm.

1.1 Early Mathematical Foundations

Grammatical transformations are inspired by mathematical transformations which are prevalent in set theory, the branch of mathematics concerned with sets and the relationships

\(^1\) The definition of terms that appear in bold type can be found in the appendix on page 1
sets of numbers with a given property are studied using the concepts of set theory. However, any set can be studied in set theory, including sets of phrases, clauses, and \textbf{lexical items}. An essential concept of set theory is the mathematical transformation which maps members of one set to another set (Moschovakis 1994, p. 2-3). For example, the function $f(x) = 2x$, where $x$ is contained in the set of all integers, is a mathematical transformation that maps the set of all integers to the set of all even integers.

Throughout his theory of generative grammar, Chomsky uses ideas clearly based on set theory. He states that the goal of syntactic theory is to explain the methods used to define the set of all well-formed sentences. Mathematical transformations map elements from one set to another set. We explore below how Chomsky invoked this idea below when developing his theory of generative grammar.

Even before the inception of generative grammar, mathematics has lent insight to linguistic theory. For centuries, linguists and philosophers have been contemplating the mathematical foundations of language. In \textit{Cartesian Linguistics} (Chomsky 1966) noted that the creation of an infinite set of utterances from a finite source of vocabulary elements and structural principles (grammar rules) has been the driving force behind our analytic perception of language since it was deduced by Humboldt in the late 18th century that “the domain of language is infinite and boundless” (p.20). Humboldt asserted that speakers “make infinite use of finite means” (Chomsky 2006, p. 15), thus a language has to be a finite set of grammar rules and lexical items that can produce an unbounded set of utterances. As Chomsky (1966, p. 19-22) further discussed, Descartes had earlier observed in the seventeenth century that humans cannot simply
be reusing pre-generated language because they are capable of producing new utterances that they have never heard before, and possibly have never been previously produced, in new situations. Descartes also points out that even at the lowest intelligence levels, humans develop language. This is a uniquely human trait that not even the most intelligent non-human animal possesses (Chomsky 2006, p. 9). And the goal of linguistic theory as laid out in the generative paradigm is to account for this uniquely human trait.

Thus, Chomsky (1955, p. 82-82 & 101-103) points to explanatory adequacy as a goal of linguistic theory rather than descriptive adequacy. Descriptive adequacy only accounts for patterns without shedding any light on qualities of human language in general. An example of a statement that fulfills descriptive adequacy is that adjectives are the class of words that “er” and “est” can be attached to. Although this description does sufficiently account for a given class of words in English accurately, it does not describe human language in general. The more desirable option is explanatory adequacy which will account for human language in general, including the intuitions of speakers. Linguistic theory is considered “successful if it manages to explicate and give formal justification and support for our strong intuitions about linguistic form within the framework of an integrated systematic and internally motivated theory” (p. 101). Both are essential for linguistics.

Chomsky (1995, p. 3-4) describes the primary goals of linguistics, as to clearly define the structure of language along with the method used to derive “free expressions,” as well as how speaker/hearers understand this structure and derivation. In order to achieve descriptive adequacy, the theory of a language must account for structure and derivation of speaker/hearers. And to satisfy explanatory adequacy, an initial state must be described along with the course
taken resulting in complete knowledge of a language. The major question is how it is possible for humans to develop any of the world’s apparently divergent languages. A step toward formulating an answer to this question is to adopt a working assumption that all human languages must have common traits. Pre-generative American linguistics failed to account for these issues fundamentally because it did not consider these as questions to answer.

Inspired by the Port-Royal Grammarians, Descartes, and Humboldt, Chomsky applies many mathematical concepts to syntactic theory in order to achieve a level of explicitness and precision required by the theoretical goals of generative grammar. A generative grammar is a set of rules that provides structural descriptions for utterances. The core claim is that every speaker of every language has mastered a generative grammar in order to produce and understand utterances, and to distinguish between those that are part of the language and those that are not. Speakers produce utterances subconsciously, and may have beliefs about their knowledge of language that do not reflect the actual (if “unconscious”) process of language production. Generative grammar is concerned with the specification of the actual knowledge of language each speaker has, which may contrast with conscious beliefs about language knowledge that s/he holds (Chomsky 1965, p. 8).

Every child, regardless of intelligence, develops knowledge of at least one language extensive enough to produce and understand an infinite number of utterances, and the development of this knowledge progresses in the same way for every child. This development hinges on primary linguistic data (PLD), which can be erroneous and are surely incomplete, since the set of possible sentences is infinite. Despite this “poverty of the stimulus (POS),” children have the ability to develop a grammar for any language and are not genetically
predisposed to learning only the language of their ancestry. Generative grammar is driven by every child’s innate ability to develop adequate linguistic knowledge of any language. It is this combination that led to the operative conclusion that children are born with the innate ability to develop a language (Chomsky 1965, p. 25-26)\(^2\).

Chomsky (1955, p. 33 and Chomsky 2006, p. 172) criticized inductive methodology in linguistics noting that deduction is a stronger scientific method for discovery. Instead of looking at an array of linguistic data and developing a theory about the behavior of the restricted array, Chomsky sought to develop a unified theory that would apply to all languages. Such a theory could then be tested against the value of the prediction about languages’ grammars that it would make. It is likely that the goal of Chomsky -- to develop a deductive theory -- drove him to integrate mathematics, which is strictly deductive, into syntax.

In an effort to explain the apparent relatedness of imperative, declarative, and interrogative sentences, including both active and passive constructions, Chomsky turns to mathematics because these structures are sets that can respectively be mapped into one another. He was also motivated to incorporate mathematics into his theory by economy, which emphasizes less complex and more universal theories. Economy is used as a metric in all natural sciences -- the simpler the descriptive mechanism, the more it is likely to describe and plausibly, thus, it’s more falsifiable. Hence, economy, which specifically in syntax determines what theories are more likely to describe the formulation of phrase structures in the mind where the construction and conversions of syntax take place, was part of Chomsky’s early criteria and is crucial also in the current research program. Recursion, which is simply using a limited number

of simple tools over and over again rather than many more complicated tools, is an essential
detail of economy. The mind contains limited space and computational capacity, hence a theory
with simpler derivation that invokes recursion is more realistic than a more complicated theory.
As a property of language, recursion must find an optimal account in a theory of grammar, and
that is a central goal (Chomsky 1957, p. 24, 53-54). Throughout this thesis, we will explore
some different ways that this problem was approached; early on, generalized transformations
(which we explore in Chapter 2) provided the account. Chomsky achieved an economical theory
by developing grammatical transformations that apply to a phrase structure, which is constructed
from the lexicon, and relate the phrase structure from “deep structure” to “surface structure”,
which are essentially sets.

Chomsky’s paradigm shift was triggered by developments in mathematics and linguistics
that preceded it, which are explored below.

1.2 Pre Generative Syntax

1.2.0 Introduction

In order to explore the motivation behind turning to the insight of mathematics, it is
helpful to understand the state of linguistics prior to its incorporation. Each new step the theory
of syntax has taken has been influenced by the findings resulting from preceding methodology.
Our discussion will begin with the original introduction of generative grammar (Chomsky 1955,
Lees 1957), which, we noted earlier, drew on the insights of thinkers from the 17th and 18th
centuries as inspiration and sought to bring into fruition their ideas. Ideas which were
redeveloped in Chomsky (1965). It is also the case, as Tomalin (2002) notes, that Chomsky’s
introduction of some formalisms in generative grammar (in particular transformations) was itself influenced by the work of Zellig Harris in the 1940’s. Shortly after the appearance of Chomsky’s *Aspects of the Theory of Syntax* (Chomsky 1965), generative semantics explored a set of inadequacies of generative grammar, and sought to incorporate meaning in syntactic derivations fundamentally claiming that syntax derived from semantics. The response of generative grammarians (Chomsky 1970/1973) accounted for inadequacies in the model highlighted by the generative semantics. This reaction influenced Principles and Parameters (Chomsky 1981) to achieve the goals of generative grammar, explanatory and descriptive adequacy. And finally the minimalist program (Chomsky and Lasnik 1993) was influenced by all of the preceding theories.

1.2.1 Carnap

In the 1930’s, the German philosopher Rudolf Carnap sought to formulate a logical system to describe language used in logical arguments that was not explicitly defined in terms of logic. By doing this, Carnap brought language and logic closer than they had ever been. Rules of inference were used to construct formulae, which he calls sentences, and which are constructed from primitive symbols of the system of language. Carnap distinguished formal language from natural language, noting that “formal” implies a separation of form and meaning. He turns to natural language to explain his reasoning. Carnap reflects on the sentence *Pirots karulize elatically*. He notes that although, there is no meaning to the words in this sentence, it can be analyzed as noun + verb + adverb. *Pirots* is interpreted as a possible English noun because it is in the sentence-initial position and its final *s* can be interpreted as the plural suffix,
hence it conforms to a pattern that nouns and only nouns typically follow in English. *karulize* is interpreted as a verb because of its position following the sentence initial noun and the presence of what is interpretable as the suffix *ize* which is recognizably used with verbs. Finally, *elaticaly* can be interpreted as an adverb because it follows a verb and contains the suffixes *al* and *ly*; *al* converts nouns to adjectives, and *ly* converts adjectives to adverbs (Tomalin 2002, p. 834).

A native speaker uses knowledge of English syntax and morphology to recognize the grammatical categories that the sentence’s constituents belong to and to determine the wellformedness of the sentence, despite the apparent absence of meaning of the sentence and the speakers unfamiliarity with the words, which are not actually English words. Thus, sentences can be analyzed disregarding the meaning of the individual words based on form alone (Tomalin 2002, p. 842). Chomsky would be heavily inspired by the ability of native speakers to make grammaticality judgments on sentences that had no meaning, as we know from his famous *Colorless green ideas sleep furiously*, which we will discuss in chapter 2.

1.2.2 Zellig Harris

Grammaticality was established as a property of a system that, once established, generates (i.e., defines) all well-formed formulae in the system. This type of closed system was attractive to linguists because it could lend insight into the theory of language by providing a model of the processes involved in phrase structure description. Phrase structure formation must

---

3 See also Chomsky (1963) for further discussion.
occur in the limited space of the mind/brain, and a model which utilizes recursion to generate all possible constituents of the system was an attractive idea to linguists.

A movement towards a formalized syntactic theory was underway. Zellig Harris (1946, p. 161-162) provided a model of the processes of phrase formation. The first step in his method of phrase structure formation is to develop a list of all morphemes of a language. The operation proposed to explain phrase structure formation is substitution. In a sentence such as Why did the puppy run away?, puppy and boy can fill the same position. If a word A can be replaced by another word B in the environment C – D, then A and B are in the same substitution class. In the example above, puppy and boy are in the same substitution class because semantically the verb run requires a noun that can conceivably run in Why did the – run away?.

Substitution classes can be thought of mathematically as sets since they are essentially sets of morphemes with the quality in common that they appear in the same environment in a sentence. Substitution classes are our first example of a linguistic set.

The next step in Harris’s enterprise was to establish all substitution classes consisting of single morphemes. Sets of all morphemes that occur in the same environments are established with special subsets for morphemes that can replace each other but occur at different places in the phrase. Some complications arise when separating morphemes into substitution classes due to the influence of the lexicon in syntax. In a case related to Carnap’s observation of native speakers’ ability to label the grammatical class of nonsensical words in a nonsensical sentence, Harris notes that words such as poem and house can substitute for each other in some cases like It is a beautiful - . but not in other cases such as He wrote a whole - . and We wired the whole - . In this and similar cases, the morphemes are considered as members of the same substitution
class, even though poem is semantically linked to write and house is semantically tied to wire. However, Chomsky (1965) responded with the sentence Colorless green ideas sleep furiously, to illustrate that structure exists without meaning and semantic relations. From this perspective, sentences like He wrote a whole house are perfectly acceptable syntactically despite their apparent lack of “real world” meaning. (This complication is reflected again in the contributions of generative semanticists, which we will explore in chapters 4 and 5, and the issues they uncovered with the more formal theory of generative semantics.) Some morphemes may differ in the affixes they can be merged with. Some words can occur with the prefix un and others occur with dis, a lexical variation, while both of these sets of words can occur before ing. Harris argues that morphemes facing this complication can be considered to be of the same substitution class occurring before ing and can be arranged into subclasses according their appearance following un and dis. Morphemes can occur in more than one class, and different substitution classes are established for positions that can incorporate some but not all morphemes that can be substituted in another situation (Harris 1946, p. 163-164).

Harris (1946, p.166) limits the scope of the substitution operation because as it is defined, it can substitute any morpheme for any other morpheme in any sequence. This will clearly lead to the production of ungrammatical sequences. Thus, substitution can only operate on morphemes which are members of a substitution class that produces well-formed outputs regardless of the manipulations inflicted on the sequence.

Harris’s substitution can be thought of as a mathematical transformation because it maps constituents from substitution classes, which we’ve seen are sets, onto well-formed structure, another set. And the limiting of the operation substitution to only apply to substitution classes
that will result in well-formed outputs is a restriction on the domain of the mathematical transformation. In mathematics, a domain is essentially the set on which a mathematical transformation can apply. In order to preserve the output, the set of inputs is restricted to only allow constituents that will form an acceptable output.

Harris (1946, p. 167-170) separates English morphemes into more than thirty substitution classes, each defined by their relation to other parts of speech. For example, the substitution class for nouns is defined as morphemes appearing before the plural affix, after the, and after adjectives. Then, the acceptable sequences of substitution classes that form grammatical phrases are defined. This was the classical American Structuralist “slot and filter” grammar.

Harris (1946, p. 178-182) noted that the utterance formulae provide for utterances that speakers do not utilize because this method does not account for limitations of speaker selection. The formulae also do not give information about the meanings correlated with the individual substitution classes and the distribution of a given morpheme. And he defines syntax as the formulae for the acceptable sequences of substitution classes that produce grammatical outputs in a given language. Harris’ goal was formalism. He was not motivated by explanatory and descriptive adequacy, so it was not problematic to him that this approach over generated.

Clearly inspired by set theory, Harris separates morphemes into sets that are compatible with each other in certain situations, and then defines formulae which describe the acceptable occurrences of the morpheme sets. He restricts set inclusion by specific formally defined traits. His next step is to incorporate the idea of transformations into his theory of speakers’ linguistic knowledge.
While setting up equivalence classes of sentences in order to study connected speech, Harris (1952) introduced grammatical transformations and was the first to use transformation in reference to syntax (Chomsky 1995, pg. 23-24). He proposed that although some sentences have different structures, they may be part of the same equivalence classes.

Harris (1952) further defined equivalence of elements as the ability to occur in the same environment in a sentence, and he defined equivalence of sentences as the ability to occur in the same context in a text or language. Also, for Harris, two sentences are grammatically equivalent if they contain the same morphemes except for differences in syntactic form (e.g. active and passive). Although grammatical equivalences have altered form, they sustain the morphemes and their grammatical relations to each other. Harris calls the active and passive forms of sentences inverses, and he noted that in addition to all other grammatical equivalences they are transformed from one form to another by grammatical transformations. For example, a sentence of the form N₁ V N₂ can have a grammatical transformation applied to it and the result will be a phrase structure of the form N₂ V* N₁. An example is the grammatical transformation of a sentence like *Abe walked the dog* into *The dog was walked by Abe*. V, *walked*, is converted from the active form to the “passive” form *was walked*, V*, yet the sentences are considered equivalent by Harris’ description (Chomsky 1995, pg. 23-24).

Although Harris’ idea of grammatical transformations based on mathematical transformations influenced Chomsky’s theory of generative grammar, Chomsky’s grammatical transformations have a different application and domain. Harris’ transformations mapped sentences with more complicated phrase structures, such as the passive and cleft structure (e.g.
The dog was walked by Abe to It was the dog that was walked by Abe), to the more basic active forms in order to explore the relationships among sentences in discourse (Chomsky 1955, p. 42).

Chomsky (1955, p. 42-44) noted that Harris’ grammatical transformations were used in an inductive way. They define equivalence classes of sentences based on observed relationships among sentences in discourse. In contrast, Chomsky’s grammatical transformations are used deductively in phrase structure description mapping structural descriptions to other structural descriptions. The next step in this exploration is to turn to Chomsky’s utilization of the insights of these linguists, mathematicians, and philosophers in the formulation of generative grammar, which we will do in detail in chapter 2.

1.3 Mathematics in Syntax - A Brief Summary

Chomsky (1955) noticed that syntax could be axiomatized utilizing the concepts of set theory. A theory of syntax that incorporates mathematics is more economical in that one set of movement and formation operations can be used to explain all movement and formation situations. The human brain has a limited capacity, and a theory that utilizes fewer rules and more general linguistic knowledge is a more realistic theory. With finite resources, an infinite language can be produced.

Importantly, generative syntax employs set theory, which provides for the notion of set, a collections of objects. Usually, sets of numbers with a given property in common are studied using the concepts of set theory. However any set can be studied in relation to set theory, including sets of phrases, clauses, or lexical items. Sets contain elements, and are determined by these members.
These concepts formed the basis of generative grammar which sought to explain phrase formation in the most economical way possible. Mathematical transformations then lend insight into the process that developed Universal Grammar (UG), the linguistic knowledge every baby is born with, into a language specific grammar. And it was the introduction of the questions (1955 and 1965) that introduced the generative framework that provided for Universal Grammar (UG).
2.0 Introductory Remarks

Generative grammar and the revolutionary innovations accompanying it dramatically changed the field of linguistics and marked the beginning of the theory of syntax. The research methods used in hypothesis development and the logic used in the study of language, most dramatically in syntax, shifted from inductive to deductive as a result of the paradigm shift. Because Chomsky sought to develop a deductive theory of syntax, he was attracted to the deductive nature of mathematics in general and the concept of mathematical transformations in particular. Generative grammar, founded in Humboldt’s ideas, recognized and attempted to account for the infinite capacity of language, its mental representation in the mind/brains of speakers, and the requirement of more than one level of representation to capture the knowledge that speakers must have.

2.1 Motivation

In order to study language in a new and more insightful framework, with the potential to understand language and language users on a deeper level, Chomsky (1955) turned to the idea of an ideal speaker/hearer conceived to have a complete understanding of a language in a homogenous speech community and be (in this ideal state) unburdened by such things as memory capacity, attention span limitations, and “erroneous” usage (i.e. mistakes by speakers such as false starts and incorrect usage) (Chomsky 1965, p. 3-6). Memory limitations, among other physiological factors, disrupt the language of real speaker/hearers. Thus the ideal speaker/
hearer as a representation of a speech community does not exist. Any one speaker/hearer can only represent him/herself, and no one is an ideal speaker/hearer insofar as no individual speaker is a “perfect” representation of a speech community. Thus the inclusion of the concept of the ideal speaker/hearer allows linguists to study language on a more meaningful level, uncovering its full potential and is not restricted by the imperfections of the users of what may thought of as be a perfect system.

Chomsky sought to account for the ideal speaker/hearer’s ability to produce and comprehend an infinite number of sentences with which they have had no prior experience, and focused on speaker competence rather than performance, which can be affected, as noted above. Rather than only classifying particular examples within a limited domain of speaker performance, the innovative\(^4\) theory Chomsky developed proposed to account for the grammar of a given language, and any native speakers ability to comprehend and produce utterances in an immeasurable range of situations with which they would have had no experience was constructed from primary linguistic data (Chomsky 1965, p. 3-6). These grammars were constituents of a **Universal Grammar (UG)**\(^5\), a theory of grammar that provided for the childhood development of these language specific grammars.

Importantly, the mathematical formalism of recursion, which was introduced in chapter 1, was claimed to be a critical component of this underlying theory of grammar. The recursive nature of mathematical transformations on sets accounts for the infinite nature of language. A mathematical transformation can be applied over and over to a constituent of a set, modifying it

\(^4\) Innovative for most of the 19th and 20th centuries, but not for the 17th and 18th centuries when this type of thinking was being cultivated

\(^5\) akin to du Marais Grammaire Générale
further and further. Looking back, in Chapter 1, at the example used to illustrate the concept of a mathematical transformation, \( f(x) = 2x \) for all \( x \) such that \( x \) is a natural number, the result of any application of \( f \) can have the transformation applied to it again and again. \( f \) applied to the result \( 2x \) becomes \( f(2x) = 4x \). Then \( f(4x) = 8x \), and so on.

Although we are focusing on the mathematical foundations, it is critical to remark on the psychological innovations of generative grammar. Chomsky introduced a new way of thinking about the status of language. Native speakers are not aware of the rules and processes involved in a generative grammar, yet everyone has mastered an ability that a generative grammar accounts for. This framework involves postulating a model of this mastery, the model being a “generative grammar,” which provides for the enumeration of all and only well-formed strings in the language. Thus, the theory of generative grammar is a mentalistic theory based on the mental reality that underlies behavior\(^6\). Generative grammar is concerned with a system beyond the capacity of consciousness, and is focused on speaker competence rather than performance. Hence the theory of generative grammar bears on the actual linguistic knowledge of a speaker and is concerned with the assignment of structural descriptions to phrases rather than the mental process involved in speech production. (Chomsky 1965, p. 4-9)

The primary linguistic “data” (PLD) that a speaker is exposed to, data which may be erroneous and incomplete (because PLD refers fundamentally to the language surrounding a child and never will those data include “all” the possible sentences), is certainly limited in ways that the speakers’ knowledge of language is not. For one thing, the data children are exposed to are finite, not infinite. However rich a child’s exposure to a language is, it will never include all

\(^6\) Here mentalism is once again a 20th century innovation, but not necessarily innovative in 17th or 18th century thinking.
the well-formed utterances that a child will be able to produce and understand in the course of his or her lifetime. Such observations are referred to under the umbrella of the “poverty of the stimulus” (POS). But, amazingly, these PLD appear to be all a child seems to need to develop an internal grammar. Every child develops the linguistic knowledge necessary to express and understand complex utterances relatively quickly. If an innate mechanism used for the acquisition of language is hypothesized to be present in humans from birth, the language learning task in the context of the POS becomes manageable. The fact that any person, without tremendous disability, learns a language with ease in the earliest years of life, regardless of intelligence and the relatively consistent order of development experienced by all learners, provides additional support for the inclusion of a language acquisition device in the theory of grammar. This task of learning cannot be managed inductively precisely because it is based on incomplete data. Hence, the poverty of the stimulus (POS) argument (Chomsky 2006, p. 141-142) points to a deductive method of language acquisition beginning with an innate mechanism designed for language learning which includes innate information including a restriction on the form of possible languages, a Universal Grammar (UG) (Chomsky 2006, p. 141-142). Thus, children must be equipped with a system guiding them toward the development of a grammar based on all (that is, any, but only) possible human languages. With an internal system from which it is possible to develop a grammar for any possible human language, children develop a grammar for their native languages. Generative grammar accounts for this human ability and maintains the goal of formulating a theory which justifies the development of a grammar given exposure to only primary linguistic data (Chomsky 1965, p. 25-26).7

Linguistics thus was redefined to respond foremost to two new inquiries. The first was about the nature of speakers’ developed knowledge of specific languages (the linguistic competence, also (as of 1986) referred to as I-language (Chomsky 1965, 1986, respectively). And the second was to address the problem of language acquisition. These question have never changed, and linguists must still account for these inquiries in all propositions and analysis of language.

Early in his work, Chomsky (1955, p. 80-83) argued that the theory of grammar is scientific, and should thus use observable data to construct a falsifiable hypothesis based set of formulated general laws. This hypothesis would be a grammar, a system that would account exactly for the well-formed utterances of a language, and speaker judgments would engender the hypothesis and new judgments would test it. Hence, Chomsky sought to formulate a theory of grammar based on speaker judgments about sentences perceived well-formed by native speakers of a given language. The structure of the set of all grammatical utterances must satisfy the laws, or rules of phrase structure formation, on each linguistic level. Because language is infinite, it is not possible to observe every possible utterance; however, every scientific theory only has some of the possible observable events available at its conception. Even so, it is possible to develop a theory that generates all possible utterances of a language based on a finite set of observed utterances. And ultimately the goal is to develop a theory that provides for the set of all possible grammars, and only those that would be possible human languages.

Although developing a theory of grammar for every human language would undoubtedly lead to a generalized theory of the syntax of human language, Chomsky asserts that developing a general theory about the nature of human language is more achievable and effective. However,
neither type of theory can be developed independently, so Chomsky looked to both methods to develop a new linguistic theory (Chomsky 1955, p. 80-83).

A linguistic theory is justified by its relation to the judgments about strings of native speakers. These judgments, rather than utterances produced by speakers, are the relevant linguistic data. The theory must provide for an individual’s internalized I-language as well as be conducive to a theory that describes all human languages. For example, the definition of a syntactic category, such as the “adjective category” in English, must be applicable in some relevant way to “adjectives” in all other languages. An adjective cannot be exclusively defined as a word to which ‘er’ or ‘est’ can be attached, because this criterion is not an applicable basis of judgment for every language. Instead, there must be a universal trait common to all adjectives in every language used to define the syntactic category of adjectives (if the category specification itself is accurate). It could be the case that a grammar developed for a given language does not coincide with a theory that describes human language. Then either that specific theory of grammar or the theory of the grammar of human language would then be considered an unacceptable theory. Thus, the general theory of grammar contributes to the grammar of any given language (Chomsky 1955, p. 80-83).

Chomsky (1955, p. 84-85) argued that the goal of linguistic theory is to avoid methodological intuition (relying on mental instincts rather than empirical evidence) as a means to reach conclusions about a theory of grammar. An abstract and general unified theory of language that explains the derivation of any grammar is the goal of linguistic theory. In the development of generative grammar, Chomsky asserted that meaning should also be avoided as a tool of syntactic analysis, because meaning does not add clarity to syntax and leads to circular
arguments (Chomsky 1955, p. 87). Inspired by Carnap’s nonsensical structure argument discussed in chapter 1, Chomsky noted that meaning does not contribute to the grammaticalness of an utterance. He illustrated this point with the now widely cited sentence *Colorless green ideas sleep furiously* which is perfectly well-formed with respect to the grammatical judgments of English speakers despite its semantic anomaly. This contrasts with strings like *furiously sleep ideas green colorless* which would be wholly incompatible with the syntax of English.

Linguistics arguably then was in search of a theory that would at least define “grammaticalness”, specify a set of grammatical sentences when the theory is applied to a finite set of data, and generate grammatical structures that are consistent with native speakers’ intuitive sense of grammaticality (Chomsky 1955, p. 94-96).

One aspect of Chomsky’s motivation for developing the theory of generative grammar was to establish a method of evaluating grammars and of comparing two grammars of the same language to determine which provides a “better” account of the language that was objective and not based on intuition. The word “better” requires definition. Thus, the form of grammar, or “the empirical form of language,” as well as the empirical consequences of adopting a particular approach to language structure was the focus of the development of generative grammar (Chomsky 1957, p. 56).

In pursuit of this deductive approach, Chomsky (1955) criticizes the use of induction in linguistic research, where most linguistic descriptions preceding Chomsky (1955 and 1957) were justified through induction. Deduction, on the other hand, produces a necessary conclusion, one that can be falsified. This contrasts with induction, which can provide only contingent results. The logical superiority of deduction inspired Chomsky to develop generative grammar utilizing
deductive logic. His pursuit of a deductive theory of language is possibly what drove him to incorporate mathematical ideas in that mathematics is driven by deduction.

Chomsky (1957, p. 13) assumes that grammaticality judgments, which are themselves speaker intuitions that allow language users to determine the grammatical and ungrammatical sequences of a language, are a valid form of judgment. Thus, sequences can be assumed as grammatical or ungrammatical based on the assessment of native speakers of a language. Speaker intuitions, including judgements about ambiguity and paraphrase, not only language corpora, are the most relevant data. Note, however, that speaker intuitions of string grammaticality contrasts with the methodological intuitions of language users concerning language development or the derivation of strings, which (discussed above) Chomsky rejects as irrelevant to linguistic theories.

2.2 The Logical Structure of Linguistic Theory

In the seminal work *Logical Structure of Linguistic Theory (LSLT)* (which we have been citing), Chomsky (1955) explicitly developed a theory based on all of these theoretical prerequisites. In order to form a viable theory which accounts for the concerns above, Chomsky incorporated mathematics and posited a theoretical basis for human language on a level never approached before. We will begin exploring the the contemporary history of the generative model of syntax by examining its first work, Chomsky (1955), in order to explore the relevance of math in the theory as it developed and its application in the modern approach.
2.2.1 The Logical Structure of Linguistic Theory and mathematical foundations

Chomsky (1955) explored the requirements of a good theory of syntax, and considering the observations above, constructed the theory of generative grammar which began with LSLT. Chomsky defines the goal of a generative syntax is to parallel any speaker’s capacity to differentiate between the grammatical sequences of a language and the ungrammatical sequences which are not contained in the language, as well as to explore the structure of the language’s grammatical sequences. The grammar of a language should construct all of its grammatical sequences and not construct any of its ungrammatical sequences, thus modeling speaker knowledge.

A grammar of a language is the set of rules used in the derivation of the elements of the set of all grammatical sequences. The derivation of a sequence is the set of rules applied to the sequence and the order in which they are applied to result in the final output, another grammatical sequence. However, the order in which the grammatical rules are applied is not illustrated by the output of derivation, which contains only what is essential for phrase structure. Thus, an output can be uniquely constructed from a derivation, but multiple derivations can be constructed from a given output. Two derivations are considered equivalent if they produce the same output, and it is possible for languages to contain occurrences of structural homonymy in which the same sequence of words can be represented by more than one distinct derivational output. And example is the ambiguous sentence, Visiting relatives can be tiresome. This sentence has two distinct structures, and hence two distinct meanings (Chomsky 1957, p. 27-28).
2.2.2 Levels

Turning to the levels of syntactic structure, Chomsky is concerned with a theory of linguistic structure, which is in essence the relationship among abstract levels of phrase structure representation. Abstract levels exist so that constituents of that level may undergo the process of grammatical structure formation in the most local of contexts. In other words, these structures reveal the relationships among elements that may be obscured or unobservable. The phonetic, phonological, morphological, word, phrase structure, and transformational levels are dependent on one another to form a grammatical output. For example, elements of the morphological level, perhaps, un and happy are concatenated (illustrated with the symbol $\cap$) at the word level to form the constituent unhappy ($un \cap happy$). And the resulting unit can itself be concatenated with -ness, yielding a new constituent (Chomsky 1955, p. 105-106).

A level $L$ consists of a fixed and finite number of elements called “primes.” Two primes in $L$ can combine by concatenation to form another element in $L$. If $a$ and $b$ are primes in $L$, they can form $a \cap b$ and $b \cap a$ which are also elements of $L$. Any two elements $X$ and $Y$ of $L$, not necessarily primes, can be concatenated to form a new element in $L$, $X \cap Y$. All of the elements of $L$ are strings, and every non-prime string consists of a unique order of primes (Chomsky 1955, p. 105-106)\(^8\).

$\cap$, concatenation, is a closed binary operation which applies to the set of all primes. This operation is closed in that it produces a new member of $L$ when it is applied to two constituents of $L$. Also, $\cap$ has an identity element $U$ which produces the original constituent when the operation is applied (Chomsky 1955, p. 107). These are properties borrowed from mathematics

---

\(^8\) The beginning of “Merge.”
to describe the behavior of phrase construction in language. Plausibly the desire for a concise
and therefore revealing theory of language as a computational system led Chomsky to look to
basic principles of math for descriptive and explanatory mechanisms.

2.2.3 Levels and Sets

Exploring the structure of levels themselves, Chomsky (1955, p. 107-108) described
individual levels as consisting entirely of grammatical structures, called $L$-markers, as they
appear on that level. For example, the morphological level consists of grammatical morphemes.
An $L$-marker contains all relevant information about a structure, called a prime, on that level.
Elements are mapped among levels by grammatical transformations, and the most important
grammatical transformation is the operation that maps $L$-markers to grammatical utterances. It is
important to note that this is not necessarily a direct mapping, and can go through some
intermediate level(s) during the process of mapping to a grammatical utterance. Language is a
system of these levels along with the operation which map constituents among the levels and the
concatenations relation which expands phrase structures within a level.

Levels are in essence sets consisting of constituents of that level (i.e. all constituents must
be of the appropriate form and have the requisite properties to be included in the set.) For
example, to be included in the set of the morphological level, a constituent must be a morpheme
and have all and only the properties relevant to being a morpheme. Grammatical transformations
map constituents among levels and therefore satisfy the definition of a mathematical
transformation since they simply map constituents from one set to another. As noted above, it is
possible for the mapping operation to pass through an intermediate level in its mapping process.
This does not conflict with the definition of a mathematical transformation. In this case the mapping becomes a \textit{composition} of mathematical transformations (one for each mapping between two levels), and the composition of mathematical transformations is a mathematical transformation. Hence, the output of a transformation can the the input to a subsequent transformation.

Utterances are generated by a recursive sequence of mappings from level to level and these mappings must be mechanically recoverable in order for information contained in previous levels to be available to a any other level. Recursion is an economical tool used in this theory of grammar. Because the same element or information can be used again and again, few elements have to be stored (Chomsky 1955, p. 114).

We have established the structure of individual levels and the mappings between them and looked at the relevance of mathematics on levels. Now, the next step is to explore the system as a whole, i.e the grammar.

2.3 Grammar

The complexity of a grammar is reduced further if a recurring set of rules can be developed which apply in a specific order for every utterance. If the order does not have to be specified in every derivation for this set, even more economy, and greater generality, is gained. Thus, a grammar meets optimal conditions when all derivations can be formed by going through a sequence of rules from beginning to end: no conversion appears twice in the sequence, and the conditioning context of each rule specifies the exact relevance and environment in which it applies (Chomsky 1955, p. 125).
Chomsky (1957, p 29-33) defined every grammar as being a finite set of initial strings and a set of the instructional rules used in derivation. Terminal strings are grammatical sequences on which the rules cannot apply any further. Thus, the terminal strings are governed by the grammatical transformations of the language.

On the level of phrase structure, where the primes are sentences, individual words, and grammatically functioning morphemes, the recursive nature of grammatical transformations can be illustrated. It is possible that an NP can be analyzed as containing another NP, such as the sentence \([NP \text{The cat [that ate [NP the mouse]]]} \text{ is orange}\). This sentence can be analyzed by the following conversion:

\[
\begin{align*}
\text{NP} & \rightarrow \text{NP}_1 \cap [\text{that} \cap \text{VP}] \\
\text{VP} & \rightarrow \text{V} \cap \text{NP}_2
\end{align*}
\]

where [ ] is an element, a prime, which prevents the formation of phrases that are not well formed. Recursive conversions like this allow for an infinite language to be numerated by a finite number of elements (Chomsky 1955, p. 172). And they also promote economy because \textit{the cat that ate the mouse} can be analyzed as one NP rather than all of the individual constituents. By using recursive phrase structure rules in this type of structure, the mind/brain gets “the most bang for the buck,” so to speak, in that these individual rules can be applied as many times as needed to create a grammatical structure. This aspect of grammar eliminates the need to store a large or infinite number of rules to create grammatical structures.
2.4 Motivation for Transformations

One of the issues motivating the development of a grammar is any speaker’s ability to compute more than one meaning for certain sequences, and thus to identify certain strings as ambiguous. Thus, the sentence *The chickens are too hot to eat* can be interpreted with two meanings. The dead, cooked chicken dish, the chickens, may have just come out of the oven and could be too hot for people to eat, or it could be that in response to the sweltering day, the chickens have lost their appetites (i.e. the chickens are too hot for anyone to eat versus it is too hot for the chickens to eat). In other words, there are essentially two issues: which NP, either chickens or some unspecified NP(*people*), is the subject of *to eat*, whose subject is not pronounced, and whether chickens is to be interpreted as the subject of “*to eat*” or as its (also unpronounced) object. Chomsky (1955, p 295) notes that sentences of this sort must have multiple phrase structures to account for these speaker judgments, which correspond to structural ambiguity. The phrase structure grammar developed thus far does not account for multiple interpretations of ambiguous sentence with more than one possible phrase structure. And because speakers have the ability to use and interpret unspoken information, a grammar which is a model of speaker knowledge must have as a goal to be able to represent such things.

In addition to being able to read more than one meaning from what seems to be a single linear string, speakers can also show that they understand two sentences with approximately the same meaning using paraphrase and synonymy. The structure of passive sentences and their structural relationships to their active counterparts reflected in speakers’ interpretations of them as paraphrases were at the time unaccounted for issues. The structural relationships between active and passive sentences as well as declarative and interrogative sentences were not
explained by the theory developed up to this point in Chomsky (1955). These sentence types are presumably related to each other because speakers can connect them, and the theory should lend insight into their similarities and differences.

The theory of grammar developed thus far also had no way of defining the mapping from levels to grammatical utterances. There was no means to take the constituent structure of a string into account when performing more conversions on it. Chomsky notes that the most natural solution would be deriving the more complex sentences from simpler sentences by using grammatical transformations. (Chomsky 1955, p. 305)

In other words, there was no way to connect the “what” in *I wonder what she put on the shelf* to the missing phrase *the bust of Plato* in *She put the bust of Plato on the shelf* even though the *what* corresponds to the NP in the VP constituent. Yet, this is knowledge that all language speakers possess, and must be accounted for. Chomsky (1955) introduced movement to account for and help define the transformation $\Phi$ which maintains these connections.

### 2.5 Defining Transformations

The movement process behaves as a sequence of mathematical transformations, mapping phrases and clauses from the set of deep structures, what Chomsky labeled the “kernel”, to the set of all surface structures, i.e grammatical utterances. The kernel contains basic grammatical phrases. Mathematical transformations map from one set to another set, a mapping which corresponds to Chomsky’s grammatical transformations moving phrases and clauses from the set of all deep structures to the set of all surface structures. These grammatical transformations apply operations to sentences in the kernel, and the result is a grammatical surface phrase structure. The grammar of a given language is the set of all of the different grammatical
transformations and combinations, or compositions, of grammatical transformations that result in grammatical phrase structures.

Chomsky (1955, p. 306) introduces the level $T$ of grammatical transformations to represent speaker knowledge of these and other derivations economically. $T$ assigns $T$-markers to strings. $T$-markers note a sentence’s transformational history and explain how a sentence was derived from the kernel, the set of all phrase structures on which only the relations of levels other than $T$ have been applied. Hence $T$ satisfies the definition of a mathematical transformation by mapping phrase structures from the kernel to a set of phrase structures with $T$-markers.

In the LSLT framework the level $T$ enables any given sentence to be represented as the sequence of grammatical transformations utilized to derive the sentence from the kernel in addition to the representation of sentences as phonemes, words, syntactic categories, and strings of phrases. $T$-markers offer a more explanatory way of representing sentences, i.e as the sequence of grammatical transformations used to derive them, which explains the ambiguity of seemingly the same sentences, because they look and sound the same on the surface, derived from the same kernel sentence. (Chomsky 1955, 306-307).

Thus, $T$-markers reveal relationships among elements that are not immediately perceptible from the order of elements. Recall that in the string *The chickens are too hot to eat*, it is not possible to show how speakers find ambiguity without showing how *chickens* can either be part of the VP *to eat the chickens* or not (i.e. being outside of it, as the subject *the chickens are too hot to eat (something)*).

Chomsky (1955, p. 309-311) further observed that some sentence pairs such as active and passive voice can be related to each other and differ only by a formal property. He asserts that
this relationship can be captured through grammatical transformations. Chomsky noted that grammatical transformations are required to be clearly defined, single-valued mappings which map strings from the set of phrase structures to the set of phrase structures, and the product of a mapping must be unique. Hence, the definition of a grammatical transformation satisfies the definition of an automorphic mathematical transformation in that strings are mapped from the set phrase structures to the set of phrase structures in a strictly defined mathematical way during which the property of belonging to the set of phrase structures is preserved.

In this early model, not all grammatical transformations were obligatory for every derivation, which was a weakness because language users would be carting around extra knowledge in the form of grammatical transformations. In this case, the set of grammatical transformations is big and contains rules not used in every derivation, or perhaps used very seldomly. And because the limitations of the mind/brain are a factor, a model with many seldom used rules is not very strong. There are optional grammatical transformations. For example, the transformation that applies inflection to verbs must be applied to every derivation, but the passive transformation does not have to be applied in the derivation of every phrase structure. Thus, the former grammatical transformation is an obligatory transformation, and the latter is an optional transformation. The kernel of a grammar $G$ is the set of phrase structures on which only the obligatory transformations have been applied. Every sentence of a language will either be contained in the kernel or derived from an element contained in the kernel by optional transformations. (Chomsky 1957, p. 45)

In order to prevent overgeneration, the prevention of which is crucial to the theory of language, because speakers can distinguish between well- and non-well-formed strings,
Chomsky (1955, p. 311) proposed conditions on the set of all grammatical transformations. The domain of each grammatical transformation is restricted to strings of a certain type, which must be finitely characterized. The mathematical idea of the domain of a mathematical transformation is overtly used in these conditions.

Every grammatical transformation is made up of a set of elementary grammatical transformations which apply to the string being manipulated with the end result of the entire transformation. The string K, which the grammatical transformation T is applied to, can be broken into a termed progression from the original K to the resulting string T(Z, K). This progression is of the form \((Y_1, \ldots, Y_r)\). The elementary grammatical transformations are each part of the larger grammatical transformation, each Y completes one part of the entire operation (Chomsky 1955, p. 324). As smaller constituents of a grammatical transformation, elementary transformations also satisfy the definition of mathematical transformations. They map a string from one set to another, or possibly from a given set into that same set, and are therefore consistent with the definition of a mathematical transformation.

In mathematics, the composition of two mathematical transformations is itself a mathematical transformation leading to the conclusion that the composition of an infinite number of mathematical transformations is a mathematical transformation. Thus, the composition of multiple elementary transformations satisfies the definition of a mathematical transformation, which is good because grammatical transformations, which are compositions of elementary transformations, have already been proposed to satisfy the definition of a mathematical transformation. Using the composition of multiple elementary transformations to form a
grammatical transformation is another way in which Chomsky utilizes the economical nature of recursion to formulate an economical and therefore feasible theory.

The goal of this theory of syntax is to describe the construction and apparent complexity of a language in a way that more elegantly parallels/represents speaker knowledge than previously achieved. Elegance is essential to the study of language because language itself is an elegant cognitive process. “Elegance” is also at the core of scientific explanation in general in that it reflects the economy achieved, which was illustrated as motivation for these innovations in chapter 1. And Chomsky strived to explore the relationship among simplicity, elegance, and explanatory power. Language undoubtedly does not utilize any unnecessary subpart which can be done without because we assume it is invoked in real time with limited memory capacity. This elegance is achieved by invoking the concept of linguistic levels, grammatical transformations, and the relationships among them. The system of linguistic levels supplies a way to provide a simplified description of the set of extremely complicated grammatical sentences as well as to determine the bounds of this set. The system of linguistic levels seems necessary at this point in the development of syntax because a theory with only one level, perhaps the level of words, which would give us sets of strings of words rather than sets of multilevel derivations, could not accurately describe the complex grammar of language without the development of a grammar of unmanageable proportions. Many intuitions of native speakers could not be accounted for with only the word level, thus, the system of levels must be created to describe the functions of grammar (Chomsky 1955, p. 293-294).

As we have illustrated, the theory of generative grammar presented in Chomsky (1955) was mathematically based with a heavy emphasis on the role of sets and mathematical
transformations. Despite predictable flaws in the theory’s design, the use of these mathematical concepts led to an insightful and economical theory, which revolutionized the state of linguistics, altering its focus. The mathematical basis of this theory was pivotal not only to the insight it lent into speaker knowledge of language, but also to the change in syntax from an inductive science to a deductive science.

We will continue by looking at the changes made to this initial theory after it was proposed motivated by finding the most economical theory possible and problems that arose after its publication.
Chapter 3: Generative Grammar Revised

3.1 Innovations

In order to expand and develop a more explanatory theory of generative grammar, Chomsky’s (1965, p. 15-16) *Aspects* model explicitly divided the model into three components: the syntactic, phonological, and semantic components. Morphology, which was included in earlier approaches, both structuralist and generative, is absent from this stage of the development of syntax. An infinite set of abstract constituents is specified by the syntactic component, with each abstract constituent incorporating the information needed for a single interpretation of a given sentence. In this framework, the phonological component applies all of the phonological rules needed to account for the structural description to correspond to a phonetic representation modeling the speaker’s representation. It determines the phonetic form of a structural description generated by the rules and grammatical transformations in the syntactic component. The semantic component provides for the semantic interpretation of a structural description. Hence, the semantic and phonetic components are interpretive and rely on the syntactic component.

The syntactic component in this model must then determine a structure distinct from what is superficially accessible in a single string, thus a so-called deep structure. This deep structure governs semantic interpretation. And a related surface structure, which regulates phonetic interpretation must also be included in the syntactic component. In *LSLT* (Chomsky 1955), the abstract “deep structures” correspond to structures that have the discontinuous and displaced constituents of the kernel sentences described in their more “local” structural settings. They are
related to the so-called “surface structures” that are closer to the more superficially accessible strings by a set of transformations. Chomsky (1965) achieves this description by expanding the mathematical basis set in Chomsky (1955) and by continuing to utilize mathematical transformations and sets as core to the model.

Figure 1

The surface and deep structures of a given structural description for a kernel (single-clause) sentence are both generated in the syntactic component and are related by grammatical

---

transformations, which provide for surface structures. In *Aspects*, multi-clausal structures are accounted for by generalized transformations which concatenate single clauses (Chomsky 1965, p. 17). These generalized transformations satisfy the definition of a mathematical transformation because they map structures from the set of all possible phrase structures to the set of more complex well-formed phrase structures. They are also recursive, an important explanatory factor in the theory. The surface structures of a language form a set as do the deep structures of a language, and phrase structures are delivered to each set by grammatical transformations which satisfy the definition of mathematical transformations. In other words, phrase structure rules are also grammatical transformations.

The syntactic component contains a base which is “a system of rules that generate a highly restricted ... set of basic strings, each with an associated structural description called a base Phrase-marker” (Chomsky 1965, p. 17), the corresponding structural description that corresponds to a deep structure. The sequence of base Phrase-markers that underlies a structural description is the basis for the structural description. The syntactic component also contains a transformational subcomponent which maps the basis to the surface structure. The set of kernel sentences is a proper subset of the sentences with a basis of single base Phrase-markers of minimal complexity (Chomsky 1965, p. 17-19). This all follows from the mathematical ideas, discussed in Chapter 2, used to develop the generative theory of grammar. The transformational component contains elements which satisfy the definition of mathematical transformations; the base is a set to which mathematical transformations apply.

---

10 Chomsky (1995, p. 25) refers to kernel sentences as the simple active declarative sentences to which obligatory transformations apply. And Chomsky (1965, p. 18) notes that kernel sentences are single clause declarative active sentences. They are what we think of as speakers, but they are not a grammatical element.
3.2 Diverging from Earlier Approaches

One motivation for the development of this generative approach is again the need to account for speaker judgments, which include not only well-formedness, but also ambiguity and paraphrase. These had not been addressed by pre-generative grammar theories as we began to discuss in the treatment of LSLT in chapter 2. Recall the sentence *The chickens are too hot to eat* has two interpretations: ‘It is too hot for the chickens to eat’ and ‘The chickens are too hot for anyone to eat’. This pair of interpretations is accounted for by attributing two distinct abstract (“underlying” or “deep”) necessarily distinct structures each related to the respective surface string (Chomsky 1965, p. 21-22). The speaker/hearer is “aware” of all of these representations and capable of their production/interpretation, and is thus able to recognize the appropriate meaning and representation in either usage.

Another motivation for change is set of speaker intuitions about the pair of sentences *John is easy to please* and *John is eager to please* which seem identical in “form”. In fact, they are a minimal pair - apparently identical in all respects except one. However, this is a minimal difference related to a difference in structure, which is expressed by forms to which grammatical transformations have been applied. And this difference is reflected in the fact that the “easy” string corresponds to the companion *It is easy to please John*, but the “eager” string does not. That is, *It is eager to please John* is not parallel to the impersonal structure with “It is easy...”. The ‘it’ in the ‘eager’ structure cannot be non-referential; speakers are unable to impute any but a referential sense to the pronoun in this case. Given the definition of grammatical...
transformations\textsuperscript{11}, because the same grammatical transformation does not seem to apply to the abstract structural descriptions of both active sentences, they cannot be of the same structural description form, i.e. in the same set. As in the early stage outlined in Chapter 2, the theory of Generative Grammar explicitly acknowledges that surface forms do not reveal everything about a sentence that a speaker is claimed to know, and thus that deep structure is vital for an account of this speaker ability. This acknowledgement, and its claim about speaker knowledge, were major innovations in linguistics. Speakers may not be explicitly aware of the structural difference between sentences with similar surface structures; nonetheless they are sensible to them. Thus, studying surface structure alone cannot provide an accurate picture of the properties of language and, importantly, linguistic knowledge, although speakers may not be aware of the properties of language and linguistic knowledge that they posses. Hence, other forms of a phrase structure must be examined along with the grammatical transformations that result in surface structure (Chomsky 1965, p. 23-24).

3.3 Grammatical Transformations

Recall from chapter 2 that in the \textit{LSLT} framework grammatical transformations apply to the structures corresponding to the restricted set of kernel sentences by performing some essential grammatical operation, and grammatical transformations mapped clauses from the set of all deep structures to the set of all surface structures. In the subsequent framework, introduced in \textit{Aspects} (Chomsky 1965), grammatical transformations map the phrase structure basis of a sentence to its surface structure, assigning the basis a derived phrase-marker in the process. A

---

\textsuperscript{11} Recall they will apply to any designated structures in a set.
transformational marker records the transformations used by a structure in derivation and does this using an alphabet of constituents which represent different grammatical transformations. The form and function of a transformation-marker is analogous to that of a phrase-marker. The transformation-marker contains a structure’s basis and denotes its deep structure. The surface structure is derived from the phrase-marker by the application of the transformation-marker (Chomsky 1965, p. 128-132). Once again, grammatical transformations are equivalent to mathematical transformations, and the mathematical framework developed in Chomsky (1955) persists.

An innovation of the Aspects framework is its treatment of the semantic component’s relation to grammatical transformations. Grammatical transformations interrelate phrase-markers that have already been interpreted by the semantic component, and outside of this contribution, the semantic component is independent of transformations. Hence, grammatical transformations cannot introduce constituents that contribute meaning to the structural description, and the constituents deleted by grammatical transformations must be recoverable. Grammatical transformations were claimed to not alter phrase structure meaning. But this seems theoretically undesirable in pursuit of modeling speaker intuitions. For example, there are meaning differences when even some of the simplest of passive transformations are applied. When speakers use the passive form, they express a different meaning and highlight or obscure a different component of a sentence. In addition, the auxiliary fronting of interrogatives changes the locutionary status of a string, changing the meaning.

The earliest transformations, the singular transformations which we discussed in Chapter 1, did not affect meaning. But others, like those involved in *The chickens are too hot to eat*
example, must affect meaning because they obscure the meaning and structure of the respective
deep structures at the surface output, at the very least. More explicitly, two distinct sets of
grammatical transformations are applied to each deep structure form and manipulate two distinct
structures. The result can be identical surface outputs. Thus, the tools used to produce an
utterance are what provide the form-meaning connection that defines language.

3.3.1 Transformational Cycle

The observation that structure is so vital both to meaning and to the representation of that
meaning led to new insights on the way grammatical transformations apply to phrase structures,
in particular, the inclusion of insights on the order in which grammatical transformations must
apply. Generalized transformations, which, recall, concatenate and embed single-clausal
sentences corresponding to the kernel, were used in the model but were not themselves ordered
(Chomsky 1965, p. 132-133). But this feature modification lent inconsistency to the theory of
syntax. Some simple grammatical transformations must be the result of complex rules (which
we will see turn out to be compositions) which could not possibly be present in a grammar or
utilized by the mind/brain due to its limited computational capacity, as a single complex
grammar rule. In order to account for both semantic and generalized transformation
inconsistencies, the model for derivation was altered, the mathematical consequences of which
are discussed in the next section.

Chomsky (1965, p. 134) eliminates generalized transformations and transformation-
markers from derivational process of generative grammar. Wherever a phrase-marker has a
position in which a clause can be embedded, #S#\textsuperscript{12}, a constituent that can be replaced by a clausal phrase structure is introduced into the existing phrase structure in the position that the new phrase structure can be embedded. For example [The cat [#S#] purred] become [the cat [the mouse chased the cat] purred]. The rules of the base apply to the most deeply embedded phrase structure and when the constituent #S# is reached the rules cyclicly generate the phrase structure that replaces #S#. A generalized phrase-marker is formed by this recursive process of phrase structure derivation. The generalized phrase-marker has all of the information of the base including all information relevant to derivation and the generalized embedding transformations.

The grammar of a language contains a linear set of singular transformations which apply cyclicly to phrase-markers beginning with the most deeply embedded phrase-marker and concluding with the non-embedded phrase-marker. For example, the sentence *The man who persuaded John to be examined by a specialist was fired* includes the three sentences roughly (1) *The man was fired*, (2) *The man persuaded John*, and (3) *John was examined by a specialist*. (3) is embedded into (2) and this embedding is in turn embedded into (1). The rules of the base apply to (3) first, followed by (2) and then (1). If the cycle was not introduced, the ordering of transformations would become a paradox. In this example, the passive would have to apply twice to the same structure. Thus, the need for transformation-markers and generalized transformations is eliminated, and generalized phrase-markers are the deep structures of a grammar, which are constructed by the syntactic component. So, the kernel and generalized transformations are replaced by the cycle. This marks the “birth” of the “transformational cycle” (Chomsky 1965, p. 135).

\textsuperscript{12} Recall, this is the notation for a position in which a clause can be embedded.
3.4 The Mathematical Connection

At this point, the deep structure is lent to the semantic component to ensure its meaning and interpretability, and grammatical transformations map the deep structure to the surface structure, which enters the phonological component (Chomsky 1965, p. 135-136). The mathematical nature of this theory is seen again here in that the grammatical transformations which map phrases from deep to surface structure are versions of mathematical transformations. The mathematical connection of grammatical and mathematical transformations remains unchanged with the change in the nature of grammatical transformations, because they are still mapping constituents from set to set in a way consistent with the way they were developed. Hence, the economy and theoretical viability lent to Generative Grammar by the principles of mathematics are preserved. The introduction of the transformational cycle introduces recursion in that the same rules are used in a recursive cycle to derive phrase structures. Along with recursion comes more economy because of the restrictions on the mind/brain discussed above.

Because the goal is to account for all - but only - the well-formed sentences in a language, an essential motivation of Generative Grammar was to avoid overgeneration. Chomsky (1965) avoided this problem by restricting the domain of grammatical transformations. Grammatical transformations filter the possible phrase-markers, in that the domain of the grammatical transformation has restrictions which limit possible derivations. Each surface structure must correspond to a well-formed deep structure, i.e. only well-formed deep structures

---

13 Recall, the domain is the set on which a transformation can apply. Restricting the domain of a transformation is just limiting the inputs that the transformation can be applied to. Thus the output set is also restricted in that, there are fewer constituents entering into the transformation, resulting in fewer outputs.
are allowed in the domain of the grammatical transformation. Now, the only output of the grammatical transformation will be well-formed surface structures. Hence, the set of possible surface structures is restricted because the domain of grammatical transformation, the set of all deep structures, is restricted by the grammatical transformation and base rules (Chomsky 1965, p. 138-139). Restricting the domain of a mathematical transformation to produce only the wanted outputs is a common mathematical tool. Chomsky (1965) is clearly using mathematical ideas in the theory of Generative Grammar to produce a plausible theory exclusively accounting for grammatical sentences. Generative Grammar remains a system of mathematical transformations applying to sets, and therefore remains a mathematical theory. This mathematical foundation results in a more economical and logical theory of syntax.

In summary at this point (Chomsky 1965), a grammar includes a syntactic component, a semantic component (the nature of which has changed), and a phonological component. This component is the part of the grammar which utilizes recursive grammatical transformations (Chomsky 1965, p 141). The syntactic component consists of a base component, which generates deep structures, and a transformational component, which maps deep structure phrase markers to surface structure phrase markers (Chomsky 1965, p 141), where all of these components are mathematical sets.

The base utilizes a categorical subcomponent and the lexicon to derive deep structures, which “enter” the semantic component where they are given a semantic interpretation, and are mapped to surface structures by grammatical transformations (Chomsky 1965, p 141), which once again are analogous to mathematical transformations. The surface structure is given a phonological interpretation by the phonetic component. The base’s categorical subcomponent
contains a set of context-free rewriting rules which specify the underlying order of lexical constituents that the grammatical transformation rules can apply to and also define the system of grammatical relations that can be utilized by the semantic component to produce a semantic interpretation.

Grammatical transformational rules apply to a base Phrase-marker, and, given the cycle, apply to the most deeply embedded Phrase-marker first. The operation of the transformational component is not complete until it has applied to all of the phrase-markers in a generalized phrase-marker. A well-formed surface structure is formed when none of the transformations is blocked, thus indicating that the relevant deep structure is well-formed. The semantic information is contained in the deep structure, and the phonological information is contained in the surface structure. (Chomsky 1964, p. 143) Also, grammatical transformations whose functions had involved permutations could be eliminated from the theory of generative grammar in favor of grammatical transformations which consist of substitutions, deletions, and adjunctions. These more elementary functions provided for displacement (i.e. movement), obviating permutation, which invoked simultaneous multiple movements (Chomsky 1965, p. 144).

The changes proposed in Chomsky (1965) are consistent with the mathematical groundwork that had been laid in Chomsky (1955). Although the new and redefined sets, deep structure and surface structure, are introduced, these sets and the grammatical transformations that relate them are compatible with the definition of a mathematical transformation and the properties of set theory.

---

14 The blocking or “filtering” of phrase structure rules will become problematic, as will the structure of semantic information in deep structure.
4.0 Introductory Remarks

In the 1960’s and 1970’s, the theory of syntax moves away from its mathematical underpinnings. The model strived to explore alternatives which would more closely account for speaker judgments of grammaticality and more accurately account for semantic interpretation. In doing so syntax moved away from its mathematical foundations in an effort also to account for the two most challenging problems of the existing model. First, many linguists argued that semantic interpretation was more closely related to syntax than the theory had allowed for, having largely separated the two. Second, the model “overgenerated,” providing for strings that speakers found unacceptable, so that grammars were too powerful. Thus, there seemed to be two ways in which the existing model was not fulfilling its purpose: it was not modeling speaker intuitions - ignoring the semantic intuitions, on the one hand, and failing to reflect speaker syntactic judgments about acceptable sentences on the other. Two distinct approaches sought to account for these issues, Ross’s constraint based theory, and the slightly later meaning based theories of the generative semanticists.

4.1 Constraints: Ross’ Approach

Ross (1967) stepped away from the Chomsky based approach by exploring filters and scaled acceptability constraints in order to address issues of overgeneration. Ross (1967, p. 7) noted that Generative Grammar consists of linguistic rules which act on variables, but that the set of variables to which a rule applies to must be restricted by constraints. His work is an attempt
to more clearly define the linguistic nature of syntactic variables which had been a part of grammatical transformations since Chomsky (1955). His method of defining syntactic variables did not conform to mathematics, and was a step away from a mathematically descriptive theory.

Any grammatical transformation can be applied over a set of variables resulting in an ungrammatical output (thus, overgeneration), and the solution to this problem, Ross (1967, p. 9) argued, is to restrict the variables on which the grammatical transformation can apply. This also simplified the grammar by making the conditions of each transformation more transparent. Essentially, Ross’s goal is to restrict the domain to which grammatical transformations can apply, an idea that itself does not disregard the principles of mathematics as we discussed related to the Aspects model in chapter 3.

Requiring that a mathematical transformation be applicable to a restricted set is not beyond the principles of mathematics. A condition can be placed on the set to which the mathematical transformation is applied in order to restrict the domain of the mathematical transformation. For example, referring back to the $f(x) = 2x$ example in Chapter 1, the domain we defined was simply the set of all integers. However, we can restrict the domain further by introducing a constraint on the domain, such as reducing it to the set of all positive odd integers. So, the transformation remains the same, but the domain to which we apply it is limited. Hence, restricting the domain of grammatical transformations alone does not affect the mathematical status of the theory.

Ross (1967, p. 13) further described the flaws due first to the underuse and then the nature of constraints in the theory of Generative Grammar by exploring Chomsky’s A-over-A
The A-over-A principle was proposed in Chomsky (1963) and named by Ross (1967).
consistently account for the grammaticality judgments of speakers, illustrated by question formation and relative clause formation.

One such constraint on reordering transformations proposed does not allow constituents contained in an NP with a lexical head noun to be moved out of that NP to a higher position in the same sentence (Ross 1967, p. 127-128). Distinctions such as that made between lexical items and non-lexical items like the particle *it*, are used to preserve the consistency of the proposed constraints. Ross has simply added a condition which the constituent of the set to which the grammatical transformation applies must satisfy. The reduction of this set does not violate the principles of mathematics. Conditions can be added to transformations that relate sets to one another.

Ross (1967, p. 69-71) introduces the idea of post generation performance filters by proposing a rule in the formal grammar of languages which, instead of restricting the application of grammatical transformation, places acceptability judgments on the phrases produced by grammatical transformations and rules. Performance filters were reintroduced in Chomsky and Lasnik (1977), which we will examine in chapter 5. Other grammatical rules change structures, where the proposed rule changes only the scaled acceptability of structures. And violations of the proposed rule do not produce ungrammatical outputs. Instead violations produce a reduced acceptability which is dependent on the structure and grammatical rule to which the proposed rule is applied.

For example, both *He figured that out* and *He figured it out* are well-formed, but *He figured out that*\(^{16}\) is less acceptable. Even less acceptable is *He figured out it*. Ross proposed

\(^{16}\) Al least without contrastive stress. *He figured out THAT* may be acceptable for some speakers.
that all of these sentences should be provided for by the grammar of English, and then are evaluated on a sliding scale of acceptability.

Such scaled grammaticality does not conform to the principles of mathematics, in which an element either has or lacks a property. There is no scaled acceptability of a mathematical category; for example a positive number cannot be more or less positive than another positive number. And a mathematical transformation would not apply to a constituent that would yield an unacceptable output. Hence, Ross’s work moved the theory of syntax away from the mathematical roots of Generative Grammar.

4.1.2 Scalar Acceptability

Other constraints apply to fully formed phrase structures and eliminate them as possible utterances based on their structure as fully formed phrase structures, or the constraints reduce their grammaticality by labeling them “less grammatical” than other fully formed phrase structures. So, the measure of the quality of grammaticality is applied after phrase structure formation, and thus non-grammatical phrase structures can be produced and then eliminated after “production” (Ross 1967, p. 238-239). This definition of a constraint violates the principles of mathematics in that the concept of scaled acceptability does not exist in any capacity in mathematics. This framework also differs from the Aspects model where ill-formed deep structures would be blocked by grammatical transformations that applied to map them to surface structure; the transformations could just not apply to strings that were not part of the defined domains. In that model, if a well-formed surface structure was produced, it was derived from a well-formed deep-structure and scaled acceptability was not an issue.
The organizational tool of these constraints is the condition box which contains all “language particular constraints.” A condition box is present in the grammar of every speaker of every language, and the rules within it are understood to apply to every rule of the grammar (Ross 1967, p. 238-239). The condition box does not conform to the principles of mathematics in that its properties are not explicitly defined, as the condition box is a set of constraints that are used to judge the grammaticality of phrase structures on a scale.

Thus, the relevance of mathematics is drastically reduced by Ross (1967). Although the move to restrict of the domain of grammatical transformations remains consistent with the notion of mathematical transformation where it originated, the scaled acceptability framework and the restriction on the application of grammatical transformations to elements that will produce unacceptable or “less” acceptable outputs violate the principles of mathematics and yields a less mathematical theory.

4.2 Meaning: generative semantics

Contrasting with the theory of generative grammar, which focused on accounting for structure, generative semanticists shifted their focus to accounting for meaning and its relation to structure as the basis of an accurate linguistic theory. They claimed that meaning and structure had a more direct connection than had been accounted for by generative grammar. Katz and Fodor (1963) argued for the necessity of universality of semantic representation in order to have language-independent representations which would allow meanings to be compared across languages, thus allowing for more interesting and insightful studies of meaning and working toward the goal of understanding the nature of human language. Generative semanticists were
driven by the observation that semantic representation is an integral part of human cognition, and therefore an integral part of utterance production. This group felt that meaning, contrasting with structure, was more central to the study of language, and their theories initiated with meaning (Jackendoff, 1972, p.1). It seems intuitive that a shift away from structure and toward meaning would reduce the relevance of mathematical transformations, which are inherently structural by nature.

Katz and Postal (1964) proposed, via the Katz-Postal Hypothesis, that the underlying syntactic structure of a sentence contains all and only information relevant to phrase structure rules. Driven by the highly disputed claim proposed in Chomsky (1965) that transformations would not affect semantic interpretation, Katz and Postal concluded that deep structure is most relevant to meaning. Recall from chapter 3 that in the Aspects model, all phrase makers have been interpreted by the semantic component before grammatical transformations are applied, thus, grammatical transformations cannot introduce constituents that contribute to the already established meaning of a phrase marker. Katz and Postal agreed that passive transformations do not change meaning and meaning changing transformations such as questions, negatives, and imperatives are also meaning preserving because they have a meaning bearing component present in the sentence’s underlying structure as had been proposed as meaning changing transformations in Chomsky (1957) (Newmeyer 1986, p. 69-71/ Chomsky 1965).

---

17 Recall, deep structure at this point is the structure before any transformations have been applied.

18 Note that it was later discovered that even the passive did change meaning: "Many arrows did not hit the target" is not the same as "The target was not hit by many arrows."
This proposal paved the way for the emergence of generative semantics, which championed a theory based less around structure and focused more on meaning. At this time, paraphrase and ambiguity were thought to have a unique underlying structure. Recall, *Visiting relatives can be tiresome*, which has two distinct deep structures. The different configurations of the deep structures are what differentiates the distinct interpretations of seemingly identical surface strings. Perhaps, on the other hand were presumed to share a deep structure, and thus meaning, even as the deep structure is modified by transformations. Everything contributing to meaning was present in deep structure, hence linguists set out to explain every semantic phenomenon with syntax (Newmeyer 1986, p. 85-87).

Another major contribution which influenced generative semantics was made by the inclusion of subcategorization features of lexical items, which contribute to meaning and structure by restricting/requiring the other components in deep structure, proposed in Chomsky (1965). For example, a transitive verb requires both a subject and object, and this requirement would be fulfilled in deep structure (Newmeyer 1986, p. 88).

Fillmore (1968, p. 42) advocated the incorporation of a case system into the theory of grammar. He highlighted that every sentence has a verb and at least one noun phrase, and each noun phrase has a case based its semantic relationship with the verb. For example, in *John broke the window with a hammer, John* has the case relation agent, and *a hammer* has the case relation instrument. Also, only noun phrases with the same case relation can be concatenated. Thus, *John and Mary broke the window* is well-formed and *John and a hammer broke the window* is not well-formed. And case was distinct from grammatical function. The same grammatical
function -- subject -- could, depending on the verb, have different relationships to the event, hence, different cases: She {earned/received} her paycheck last month (Fillmore 1968, p.42-43).

Fillmore (1968, p. 49) argued that the verb creates a “case frame” which indicates all of the lexical items required to create a well-formed utterance using the selected verb and all of the case relations that the lexical items are required to fulfill. This model is, thus, heavily influenced by semantics and the semantic requirements of verbs.

Katz and Fodor (1963) proposed that semantic interpretation is accounted for by a set of projection rules. The semantic focused projection rules determine the structure and meaning of a sentence and contribute all of the semantic information needed for interpretation not contributed by the lexical items present. The result is an utterance’s syntactic structure which is determined by a sequence of phrase structure rules and transformations. Katz and Fodor argued that meaning must be integral to the theory of grammar because the only thing that differentiates two utterances with the same lexical items but distinct meanings is the rules/transformations and the order in which they are applied (Jackendoff 1972, p. 5-6). This position challenged the previously accepted idea that transformations are meaning preserving.

Katz and Postal (1964) proposed that semantic structure defined the base of utterance formation. The deep structure of an utterance was formed based on semantics and grammatical transformations applied to these semantic based deep structures to form surface structures (Jackendoff 1972, p.3,7). These ideas represented a clear shift in syntax, with meaning contributing to structure. Katz and Postal (1964) set out the fundamental principle of relating meaning to deep structure, while Katz and Fodor (1963) tried to elaborate a semantic theory that would ultimately make the Katz and Postal hypothesis work (Harris 1996, p. 81-86).
This period of development in syntax deviated from the mathematical roots of the theory. Scaled acceptability, acceptable deviations from the optimal form, and filters, which eliminated undesirable structures after their formation, violated the principles of mathematics and reduced the logical elegance of the theory of syntax. Nonetheless, the changes proposed during this period brought to the spotlight important observations about the interpretation of linguistic structure.
Chapter 5: Consequences

5.0 Introductory Remarks

Exploring further the ideas proposed by generative semantics, linguistic theory developed in two directions. The linguists that were inspired by Ross (1967) and Katz and Postal (1964) continued to explore their ideas. Generative semantics sought to change deep structure into semantic structure and then use syntax to derive surface structure. But this required scalar judgments and new varieties of interpretation because it is not the case that every constituent of a grammatical category acts the same. For example, not all verbs act the same; some have different semantic relationships to their arguments. During the same period, Chomsky and Jackendoff continued the train of thought exploring meaning relations.

5.1. Exploring meaning: Remarks on Nominalization

Chomsky (1970) attempted to refocus the “argument” back to the theoretical ideal of a context-free language, not subject to the performance errors of language users, and how language can be learnable with imperfect data. It was proposed that the syntax of a grammar includes the categorial component and lexicon, which make up the base, as well as grammatical transformation rules. Considering a theoretical ideal language rather than real world languages was a tool to develop a theory of grammar because it was easier to work with than real world languages.

Remarks on Nominalization also responded to the generative semantics focus in a particular context. Within the claim about transformations, ambiguity, and paraphrase, such relationship as those between the sentence, The enemy destroyed the city and the noun phrase the enemy’s destruction of the city, took a greater significance. Generative semanticists saw these as examples of paraphrase. And that thinking engendered Remarks on Nominalization, as the discussion of John is eager, and John’s eagerness, discussed below reflects.
language. A theory developed about a theoretical ideal language could then be compared real world data to test its viability.

The lexicon is made up of constituents with a system of lexical features, called lexical entries. In this ideal language, there is an infinite set of lexical entries. Phrase markers emerge with a “dummy” symbol, a placeholder, which contains only a set of features, in one of the nodes. This node is replaced via lexical insertion with the lexical entry whose features match the features of the dummy node. These insertions were also seen as transformations. These phrase markers make up deep structure in a model where transformations map phrase markers to phrase markers. A given set of transformations, when applied to deep structure, produces subsequent intermediate structures, and ultimately surface structure, where phonological rules are applied.

In Chomsky (1970) *Remarks on Nominalization*, the grammar includes semantic rules which would apply to each deep/surface structure pair, but only grammatical relations present at deep structure would be affected by semantic interpretation (p. 184-185).

Chomsky (1970, p. 187) reexamined the structures associated with the minimal pair *eager* and *easy* that we discussed in chapter 3. He explored the gerundive nominal, *John’s being eager to please*, and the derived nominal, *John’s eagerness to please*, both associated with the sentence *John is eager to please*. The potentially most intriguing qualities of these nominals are their internal structures, the productivity of their derivations, and “the generality of the relation between the nominal and associated preposition.”

Examining first the gerundive nominals (*John’s being eager to please*), Chomsky (1970, p. 185) noted that they are derived from subject-predicate structures which have a regular meaning relation. Also, adjectives cannot modify the gerundive and *John’s* cannot be replaced by
determiners, so gerundive nominals must not have the same internal structure as a noun phrase. Chomsky claimed that these consequences are precisely what is expected when assuming that grammatical transformations form gerundive nominals when applied to underlying “sentence-like” phrase structures.

Turning to the derived nominal (*John’s eagerness to please*), the case is completely different. The derived nominal has an internal structure parallel to a noun phrase (*The eagerness to please* is acceptable). And the production of phrases of this form is greatly restricted, in that the “semantic relations between the associated preposition and the derived nominal are quite varied and idiosyncratic” (Chomsky 1970, p. 186). Given these differences, Chomsky (1970) proposed two explanations: either the base rules are extended to include derived nominals, simplifying the transformational component (the lexicalist position), or the base rules are simplified to exclude derived nominals, and the transformational component is expanded to include rules for the derivation of these forms (the transformationalist position) (Chomsky 1970, p. 186).

Chomsky (1970) focused on the lexical hypothesis and what it could tell us about these nominals. His first step was to explore the restrictions on derived nominals, and there are some productivity issues. For example, looking at *John is easy to please*, the gerundive nominal, *John’s being easy to please*, is acceptable, however, the derived nominal, *John’s easiness to please*, is not. If the difference between the grammaticality of the two derived nominals *eagerness* and *easiness*, is approached with the transformational hypothesis, it is necessary to assign a range of meanings to the base structure and add an extensive set of conditions on semantic features of nominals stipulating which can transform to derived nominals and which
cannot. This reduces the applicability of the hypothesis that the transformational component does not contain semantic information (p. 188-189).

Recall that derived, but not gerundive, nominals have the structure of a noun phrase. Derived nominals cannot contain aspect, they do appear in the full range of noun phrase structures, and appear with the full range of determiners. The transformational approach does not account for why derived nominals act like “ordinary” noun phrases with any sort of elegance, but this property would follow directly from the lexicalist approach (Chomsky 1970, p. 189-190).

Chomsky (1970, p. 190-191) proposed the tentative hypothesis that lexical items have fixed selectional features, strict subcategorization features, and a choice as to what lexical category (noun, verb, adjective...) they can be inserted as. So, eager must contain the subcategorization feature that it can take a sentence structure as a complement John is eager [(for us) to please]. In addition, eager can appear in the noun position in John’s eagerness [(for us) to please]. But, easy on the other hand does not have this subcategorization feature in the lexicon, and thus the derived nominal John’s easiness to please is ungrammatical, and no structure of that form is generated with base structure rules, and the derived nominal in question cannot be formed. Note that this restriction is not applicable to gerundive nominals.

5.1.1 X-bar Theory

In order to capture the relationships not obvious from fully developed utterances, Chomsky (1970, p. 210) posits a system of visually representing the underlying construction of phrases in which the categorial component adheres to construction rules. Letting X be any
category (noun, verb, etc.), the following rule is proposed: \( XP \rightarrow [\text{Spec, } X'] \) \( X \) for each category present in a phrase structure. This looks like:

\[ \[
\begin{array}{c}
\text{XP} \\
\text{Spec} \\
\text{X'} \\
\text{X}
\end{array}
\]

**X-bar theory** makes evident the structural differences and similarities in phrases that go unnoticed otherwise. Seen in the derived nominal of (51) and a related phrase (52) which have similar structures (Chomsky 1970, p. 211).
The X-bar innovation became integral to the analysis of phrase structures and the development of syntax. The clarity it added to the relations of lexical items and phrase structures powers the investigation of syntax until the 2000’s.

5.2 Lexicalist vs. Transformationalist

Chomsky (1970 p. 195) argued that there was enough evidence supporting the lexicalist hypothesis for derived nominals, and so derived nominals must be noun phrases. Hence, the transformational approach had been accepted for gerundive nominals, but the lexicalist hypothesis was preferred for derived nominals. Chomsky (1970, p. 215) reflected that this mix of approaches adds confusion to the study of syntax. When both grammatical transformations AND the lexicalist approach are included, complex problems become inherent.
The transformationalist approach is consistent with set theory and mathematical transformations. The application of mathematics is essentially irrelevant to the lexicalist approach as it is outlined in Chomsky (1970). The lexicalist approach doesn’t affect the relevance of mathematics to syntax negatively, and the inclusion of derived nominals in the base has no real contribution to the mathematical approach the theory.

Jackendoff (1972) also explored the consequences of the (extreme) transformationalist position represented by generative semantics. He pointed out other problematic consequences and sought to improve the lexicalist analysis. He noted that if all meaning is present in deep structure then projection\textsuperscript{20} rules do not contribute to meaning and deep structure must consist of logical rather than syntactic structure. This leads to deep structure that is so abstract that all syntactic significance is lost and grammar becomes a collection of arbitrary restrictions on transformations. The transformational component then becomes more complex and powerful with a long list of exceptions (Jackendoff 1972 p.9).

Restricting transformations can reduce the relevance of mathematics. For example, restricting the transformation \( f(x) = 2x \) to apply only to even integers is perfectly viable. However, when the restriction becomes more arbitrary, the mathematical elegance can be lost.

Jackendoff (1972, p. 3) argued that a model driven by syntactic representation would produce “a more comprehensive and precise theory of the semantic component”. He proposed the retention of deep structure in its original sense, representing syntactic generality. He argued that his model, which included a syntax bearing deep structure, is more plausible than a purely semantic deep structure proposed in order to preserve the Katz and Postal hypothesis because the

\textsuperscript{20} See page 7 of chapter 4: A set of projection rules contribute the semantic information needed for interpretation not the lexical items present Katz and Fodor (1963).
latter did not account for all possible English utterances. Also, the Katz and Postal theory did not account for some generalizations without added constraints. And finally, the group of possible languages accounted for by Jackendoff’s theory (while still corresponding to speaker judgments) would be smaller than the set that the Katz and Postal proposal provided for. Thus, Jackendoff’s theory was already less powerful, an advantage, because a narrower set becomes closer to the goal of defining human language, which is subject to the limited memory and other constraints of the mind/brain (p. 11-12).

Jackendoff (1972 p. 18-20) proposed well-formedness conditions to restrict the semantic interpretation of possible utterances. An example of such a condition is a selectional restriction on semantic interpretation which strives to mirror a speaker’s intuitions of abnormality by marking sentences that are grammatically well formed, but have no sensible meaning unacceptable. This would mark as anomalous sentences such as *Colorless green ideas sleep furiously* and *That man is pregnant*. The incorporation of selectional restrictions allows a syntactic grammar to operate freely with acceptability being analyzed after production. But, once again, any type of elimination after formation is not mathematical. In mathematics, if a structure is impossible, or not a part of the image, that is, set of outputs, of a mathematical transformation, it will simply never occur. Basically, everything that is part of the image will be produced by the given transformation, and the entire pre-image will be included in the domain, or the set of inputs. Anything that is not part of the image, on the other hand, will not be produced by the transformation applied to domain constituents.

---

21 And later, the subject to the “learning constraints” of language acquisition.
The other main approach to solving the puzzle presented by the generative semanticists was the idea of constraints, which was also explored during this period.

5.3 Well-formedness Conditions

Lakoff (1971 p. 329-332) argued that the acceptability -- and therefore grammaticality -- of a sentence cannot be analyzed in isolation; instead a set of presuppositions must be considered. In this system, a sentence is well formed or grammatical relative to a set of presuppositions. The fact that the ability to make these judgments is part of an individual’s linguistic knowledge led Lakoff and others to maintain this position. However, the set of presuppositions involves extralinguistic knowledge in that these presuppositions will be based on an individual’s experience in the world. Thus grammaticality judgments will vary from person to person. For example, some people will judge “My cat realizes I am a lousy cook” as grammatical because they believe cats have the capacity to differentiate between good and bad food, while others will judge it to be ungrammatical because they do not believe cats have such complex thoughts.

This idea is a broadening of the notion of “grammaticalilty” to include an individual’s response to strings in the context of his or her opinions, views, or experience with the world as components of language production/understanding. This contrasts with the assertion of Chomsky (1957) that grammatical structure can exist without coherent meaning exemplified in Chomsky’s now classic example Colorless green ideas sleep furiously. Lakoff (1971) argues that it is possible to have a completely structurally sound utterance deemed ungrammatical based on an individual’s estimate of the truth value of the utterance. This analysis violates the
constraints on what a mathematical transformations can include and therefore is not mathematical. This is because mathematical transformations are strictly structural, and any information not relative to the structure of the input or output is not relevant.

5.3 Graded speaker judgments

Lakoff (1971) argued further that there is not simply a set of grammatical sentences and a set of ungrammatical sentences. But rather, speaker judgments are graded, meaning some sentences are more acceptable than others on a progressive scale. And hence, the notion that a sentence is either generated by a grammar or not possibly generated cannot be accurate. Ross (1973) also proposed a scaled inclusion of clauses in the noun category, a scale referred to as “nouniness”. A noun would have the strongest noun quality, while derived nominals, action nominals, that-clauses (that Max gave the letters to Frieda as in That Max gave the letters to Frieda was shocking), and others have the noun quality (i.e could function as a subject), but less so than lexical noun (Newmeyer 1986, p. 123-124).

This treatment of grammar as crucially including graded scales is not consistent with a mathematical framework. A mathematically relevant object either has a quality or does not, and all objects said to posses a given quality have an equal amount of that quality. For example, 2 is not more or less prime than 3; they both satisfy the definition of being prime. In this regard, the lexical hypothesis is closer to the mathematical framework with the “absolute” quality of the definition, especially if binary selectional features define the characteristic of a lexical item. Rather than use a scaled system of inclusion, the lexicalist hypothesis includes derived nominals
in the base component. Derived nominals are constituents in the set of base components, and from the perspective of set inclusion, this is a more mathematically satisfying position.

5.4 Chomsky’s Constraints

As an alternative response, Chomsky (1973) reforms the theory while still observing the principles of mathematics in a way similar to what Ross (1967) had originally proposed.

Chomsky (1973) proposed a change that was mathematically equivalent to restricting the domain of mathematical transformations. Chomsky (1973, p 232) differentiates two aspects of grammar. The first being conditions on the legitimate forms of linguistic systems that suffice as grammars, and the second being conditions on the functional rules as they are applied to structural descriptions. When applied to grammatical transformations acting on phrase markers in phrase structures, these conditions will prohibit grammatical transformations that would result in ungrammatical phrase structures, such as reversing the order of a phrase marker’s words to form the mirror image (which is a simple permutation), while, on the other hand, preserving the viability of grammatical transformations that are consistent with what speakers judge to be well-formed phrase structures, such as the “passive transformation.” Thus, grammatical transformations have conditions which must be fulfilled in order for the grammatical transformation to apply. For example, using the elementary transformation discussed above, the passive transformation will reorder the NP’s as specified by the rules of the transformation, if and only if a phrase marker is of the proper form. Hence, phrase markers such as *Perhaps John read the book intelligently* are converted to the passive form *Perhaps the book was read*.
intelligently by John while phrase markers that do not fit the specified form cannot be affected by the passive transformation.

This reformation is equivalent to restricting the domain on which grammatical transformations can act: more explicitly, the set on which grammatical transformations can apply is restricted. This model is consistent with a restriction on the domain where a mathematical transformation can apply, in that a mathematical constraint does not operate on the variables, but rather on the form of the domain set. In the above example, John is eager to please is not in the domain of the passive transformation. Because this reform mirrors restricting the domain of a mathematical transformation, the proposal can be considered a move back towards conforming toward a model consistent with the principles of mathematics, as was the earlier model.

5.5 Constraints and Filters

However, proposals made in later Chomsky and Lasnik (1977) still in the direction of limiting grammars actually move away from the principles of mathematics and the insight lent to syntax by the incorporation of mathematics. Chomsky and Lasnik (1977, p. 427) introduced the concept of optimality, which ranks the constraint violations of languages. A language will allow the most optimal construction; that is, the one violating the fewest constraints valued to be grammatical. Valuing the violations of some constraints as less critical than other constraint violations, structures which do not perfectly conform to the grammar of a language are still deemed acceptable. There exists the completely unmarked form of UG, violating no constraints, and deviations from such a formulation of UG are ranked according to acceptability in specific languages (Chomsky 1977, p. 430), another form of scaled acceptability. In the same way, the
ideas developed in Ross (1967) and Lakoff’s scalar approach to “nouniness” violate the principles of mathematics, so does this concept of optimality. Optimality ranks the violations of grammaticality allowing the least unacceptabe structure to remain grammatical. In mathematics, a constituent is never included in a set simply because it has the least unacceptable properties. Set inclusion entails complete conformity to a given property. Hence the group of acceptable utterances can no longer be thought of or studied as a set and grammatical transformations cannot be compared to mathematical transformations.

Surface structures are required to meet wellformedness conditions, which are either surface filters or interpretive rules (Chomsky and Lasnik 1977, p. 428). The filters which apply to surface structures and bar ill-formed structures take the form of ordering statements, statements of obligatoriness, and contextual dependencies (Chomsky and Lasnik 1977, p 433). However, filters reduce the relevance of mathematics to the theory of syntax because they do not conform to the principles of mathematics as illustrated above.

Filters limit the grammaticality of outputs of the system of phrase structure construction. Filters eliminate ill-formed phrase structures from the grammar of a language by declaring them ungrammatical after they are constructed due to the violation of a given filter condition. The theory of filters puts conditions on the wellformedness of surface structures which constitute the output of the operation of the functions comprising the derivational system. This, thus, diminishes the mathematical relevance present in earlier models (Chomsky and Lasnik 1977, p. 489-490), these filters violate the principles of mathematics. The system of mathematical transformations does not allow the formation of elements with undesirable traits and then eliminate them after they are formed by declaring them unacceptable. On the contrary,
mathematical transformations exclusively produce and apply to elements which conform to given specifications.
Chapter 6: Principles and Parameters

6.1: Motivation

Readdressing the initial parameters of generative grammar discussed in Chapter 2, Chomsky once again explored Universal Grammar (UG) and the theory of language that can be deduced from its consequences. Chomsky (1981, p. 7-9) reminded us that UG was claimed to represent the pre-language specific mental state of children as they begin to develop a specific language. And in one way or another, the notion has been a part of the model since Chomsky (1965).

The possible parameter settings of a language are predetermined by UG. UG contains all acceptable parameters of human language. These parameters are a set in a child’s developing - and thus language specific - language faculty. As a child develops language, he or she eliminates/ includes language specific parameters based on linguistic experience, namely exposure to the linguistic outputs of more skilled language users, such as parents (Chomsky 1981, p 7-9).

Syntax is concerned with the development from the initial state, shared by all at the common beginning of language acquisition, to the steady state, in which a grammar has been developed and changes very little and the change that does occur is predominately in the lexicon (Chomsky 1995, p.14).

Once all of the parameter settings are in place, the claim is that a speaker has developed a core grammar of a language which is aligned with (but never identical to) that developed by other speakers of the same language. Each individual’s core grammar is colored by his or her

---

22 See Chomsky 1973 pg. 232
experiences, and within limits, interpretations of the data in the context of UG. All members of a speech community will have slightly different core grammars. In the initial state, UG allows only a set of unmarked\textsuperscript{23} constructions. Based on the experiences and exposure of the language learner, marked structures can become grammatical in the developing core grammar (Chomsky 1981, p 7-9).

This core of the Principles and Parameters approach is mathematical in that beginning with a universal set of exactly the possible acceptable grammars and then restricting the language faculty further to contain information specific to a particular language is equivalent to restricting the elements included in a set. As proposed, a language learner eliminates parameters not included in a particular language as a result of exposure to primary linguistic data, which comes from interaction with experienced language users. This can be thought of as a restricting mathematical transformation mapping an individual’s linguistic knowledge from a more general set to a set that has eliminated certain unacceptable parameters.

In English, for example, nouns do not have gender as they do in other languages (such as Spanish or German). So, the inclusion of gender as a quality of nouns is an acceptable parameter of human language and included in the set of all possible parameters. But, when a child is developing English, he or she will eliminate gender of nouns from the set of parameters of English. Hence, his or her linguistic knowledge is mapped by that development from the set of parameters including nouns having gender to the set of language specific parameters that does not include the parameter of nouns having gender. This is the mathematical equivalent to

\textsuperscript{23} A construction is unmarked when it occurs more typically than a more unusual alternative. Such as the present tense vs. the gerundive. Markedness has long been a problematic category marker.
eliminating odd numbers and performing transformations with only even numbers. The system will still be functional with only evens.

A contrast is made between I-language and E-language where I-language is the Internalized language of an individual and E-language is the external supposed language of a speech community. Syntax is primarily concerned with possible I-languages, speakers’ differing core grammars as noted above. E-language consists of what essentially makes up the language of an entire speech community and does not necessarily exist (Chomsky 1995, p. 15-16).

Chomsky (1981, p.10) advocates the notion that the theory of UG is ideally a primitive set of concepts and theorems that can be used to derive the other principles of the theory in an axiomatized fashion. The emphasis on a deductive theory is parallel to the initial motivation of the Generative Grammar. The logical superiority of a deductive theory is once again emphasized. The goal of a theory that is both logically sound as well as descriptive and predictive is re-established over a theory that is simply descriptively adequate.

6.2 The Government and Binding\textsuperscript{24} model

6.2.1 Transformations

Consistent with Chomsky’s model of Generative Grammar, the model of syntactic construction consists of D-structure, S-structure, PF, LF, and the transformation component which moves structures along their paths of development (Chomsky 1981, p. 18). Once again, we have phrase structures being mapped among sets (D-structure, S-structure, PF and LF) via the

\textsuperscript{24}Probably named as a consequence of the title of the work in which it was first outlined Chomsky (1981) \textit{Lectures on Government and Binding}, and the focus on the role of government, and the relationships among the different types of nominals - lexical, anaphoric, and referential.
set of mapping rules that make up the transformational component. Hence, the theory can be examined utilizing mathematical terminology and logic.

Chomsky (1980, p.145-148) had reduced the set of grammatical transformations (which had included substitutions, deletions, and adjunctions) that map sentences and phrases to S-structure to exclusively move $\alpha$, where $\alpha$ is any phrase category. The transformational component is reduced to the single rule move $\alpha$, which relocates phrase structure constituents within a phrase structure and leaves behind traces co-indexed with the moved antecedents (Chomsky 1981, p. 5). Move $\alpha$ acts as a mathematical transformation in that it maps elements among the levels of grammar, PF, LF, S-structure, and D-structure. It also maps elements from a level into the same level, or from a set onto that same set. Thus, move $\alpha$ acts as a mathematical transformation by mapping elements among sets reflecting the model developed in Chomsky 1955.

However, Chomsky noted that this simplification would over-generate (which is once again a major problem to overcome, because a grammar of language that provides for utterances that native speakers would never find acceptable is not an adequate grammar), hence the domain of move $\alpha$ must be restricted. Restricting the domain of the grammatical transformation move $\alpha$ is consistent with mathematics because it is equivalent to restricting the domain of a mathematical transformation. In both cases, the transformation will operate in the same way, just on a smaller restricted set.

---

25 Recall, PF is phonetic form, LF is logical form, S-Structure is surface structure, and D-structure is deep structure. See definitions page.
6.2.2 Thematic Roles

Chomsky (1981, p. 35-36) took as axiomatic the semantic/thematic/case relations ideas shared by Jerrold Katz (1980), Jeffery Gruber, Ray Jackendoff (1972), and Charles Fillmore (1968). In LF, “NP’s with some sort of ‘referential function,’ including names, variables, anaphors, pronouns, but not idiom chunks or elements inserted to occupy an obligatory position of syntactic structure” are considered NP arguments and each is assigned a “theta-role”. In other words, they are “assigned the status of terms in a thematic relation.” The role played by each component of a phrase structure determines its form. Meaning, each component has grammatical features that drive its inclusion and placement, its thematic (theta ($\theta$)-role). Each position in a phrase structure exists to fulfill a $\theta$-role required by the verb, preposition, or other components, and is called a $\theta$-position. Hence, the $\theta$-criterion, which states that each $\theta$-role corresponds to exactly one argument and each argument corresponds to exactly one $\theta$-role, was introduced (Chomsky 1981, p. 36). The $\theta$-criterion restricts the formation of phrase structures, and defines well-formedness for these structures. Hence it reduces the domain of mathematical transformation move $\alpha$ and motivates the construction of phrase structures, responding, in part, to the overgeneration problem in ways still mathematically coherent.

6.2.3 Government

Chomsky (1981, p. 5) re-introduced the term “government”\textsuperscript{26} with a reference which is concerned with “the relation between the head of a constituent and categories dependent on it.”

\textsuperscript{26} Chomsky (1986, p. 162) notes that “government” is a term in grammar that dates back to the ancient Greeks and was following the approach developed in Aoun and Sportiche (1983). As we see as well, government is founded in c-command, which is attributed to Reinhart.
A constituent, A, governs another constituent, B, if A is the head of a lexical category or inflection phrase, and A does not dominate B but the maximal projection of A dominates B (A c-commands B). If there is another potential governor of B, call it C, that is not a governor of A, i.e. between A and B, then C is a barrier and A does not govern B (Chomsky 1981, p. 163).

Case must be assigned to all NPs with phonetic representation. Typically, a governing category assigns case to an NP. NPs move to get case, which is typically assigned by VPs and IPs (inflection phrases) (49-51). A binds B if A c-commands B and A and B are coindexed (they corefer). Bound constituents must appear in the same clause eliminating the coreference of cross-clausal elements, which native speakers judge ungrammatical (p.184). These can all be seen as “pieces” of an overall way of limiting the domain to which the function move α can apply. That is, the outputs will reflect the set of strings/structures that speakers of a language judge as well-formed.

6.2.4 Empty Categories: Trace and PRO

The phonetically empty node left behind by move α contributes to the structure of a string. Thus, trace is introduced to account for this. Trace is defined as a constituent without any phonetic representation left behind after move α has been applied to a constituent of a phrase structure. Thus, trace is phonetically null element that allows the position of a moved constituent and its maximal phrasal projection to remain in the phrase structure after the moved constituent has left (Chomsky 1981, p. 146-149 and 158-163). This can be seen in John seems [t’ to have

27 Case is the result of Fillmore’s influence. The thematic roles that constituents fill are motivated by case which is the grammatical function a constituent fills, and case is independently determined syntactically (i.e tense is governed by V or P). For example, the subject of a phrase/clause is typically given nominative case (I ate the apple).
The theta-role of a moved constituent is thus preserved both with the trace and the moved element (Chomsky 1981, p. 116). This property remains in all sets regardless of the application of the mathematical transformation \textit{move} \(\alpha\), thus, the mathematical transformation preserves this property and it is required in all sets among which mapping can occur.

Chomsky (1981, p. 321-322 and 1986, p. 163) distinguished between traces left by extraction and phonetically null constituents. Thus, alongside trace, another element \textsc{PRO} is introduced exemplified in \textit{they tried[PRO to be happy]}. Although also phonetically empty, it contains the features number, gender, person, but is ungoverned, where trace is governed, bounded, and transmits its theta-role. \textsc{PRO}s exist because they are ungoverned, unbounded, but needed for theta-role satisfaction and identity, i.e. speakers “fill them in.” Chomsky (1981, p. 323) argued that trace and \textsc{PRO} are actually different types of occurrences of the same empty category. They are two sides of the same coin and occur in complementary distribution. Although they behave differently in terms of government and binding, they can be thought of as the same thing. The empty category may contain values for person, number, case, gender and others, but these attributes are not required.

6.3 \textit{Movement and Case}

Movement cannot be spontaneous and must be syntactically motivated. Thus, movement is driven by the structures specified by lexical items in that constituents are moved to fulfill the case requirements of phrase structures. For example, \textsc{CompP} and \textsc{VP}\textsuperscript{28} assign case to a

\textsuperscript{28} \textsc{CompP} is complementizer phrase, or a phrase headed by a complementizer, such as \textit{that}, \textit{if}, or \textit{for}. \textsc{VP} is verb phrase, or a phrase headed by a verb.
constituent, and an NP or DP must be assigned case to receive a theta-role. An element lower in
the phrase structure is moved to the specifier position of these phrases when case must be
assigned. A constituent has case if it is in a position in which case is assigned, and a chain has
case if any member of the chain has been assigned case. Every NP must be assigned case or be
part of a chain that has been assigned case, where a chain is a sequence of constituents to which
move α has been applied with all but the initial element being a trace of that element (Chomsky
1981, p. 334). In the example, John seems [t’ to have been expected [t to leave]] has the chain
(John, t’, t) with John being the head of the chain (Chomsky 1995, p. 44). The motivation of
mapping from D-structure to S-structure is revealed to be case assignment. The mathematical
transformation move α accomplishes the repositioning of elements to positions with proper case
assignment mathematically in that it maps elements from the set where they are without proper
case assignment to the set with proper case assignment, S-structure.

Chomsky (1986, p 4) further distinguishes two types of movement: substitution and
adjunction. Substitution can only be applied when the following domain restricting properties
are met: nothing can be moved to a complement position; only maximal projections (i.e. full
phrases) can be moved to specifier or head positions; and exclusively minimal and maximal
projections can have the rule move α applied. Adjunction is restricted to be possible only when it
moves a maximal projection which contains a non-argument (p. 6). Thus, the domain of the
transformational component, move α, which still satisfies the definition of a mathematical
transformation is restricted further.

The Projection Principle is also introduced, which states that the structure of a language is
determined by the lexicon in that the representations of phrase structures at each level of
syntactic representation are projected by the lexicon. The structure S consists of a NP-INFL-VP pattern (Chomsky 1981, p. 29). Thus, the items selected from the lexicon determine the application of the mathematical transformation move $\alpha$, and, moreover, the sets of possible phrase structures in both D-structure and S-structure are reduced to conform to the “allowed” form of S.

This model (early on informally referred to as “Government and Binding”, but later more precisely defined as the Principles and Parameters model), conforms to the principles of mathematics, reflecting the motivation and roots of the initial theory of Generative Grammar. The units of grammar, S-structure, D-structure, LF, and PF, are defined by given properties to which the members of each level must conform, corresponding to well-defined sets with unique constituents. The transformational component, consisting of move $\alpha$, maps elements among levels or into the same level, satisfying the definition of a mathematical transformation. Therefore, the model of grammar is once again reduced to the rules of sets which are related via a mathematical transformation, conforming to the properties of set theory which influenced the theory of Generative Grammar developed in the 1950’s.

\footnote{The Projection Principle was very strong and caused problems for morphology.}
Chapter 7: Minimalist Program

7.0 Introduction

The Minimalist Program draws from the successes and failures of the constantly evolving theory of syntax from generative grammar to Principles and Parameters. The Minimalist Program simplifies many aspects of the Principles and Parameters model, such as movement and phrase structure, and we will explore how the Minimalist Program re-adopts the stricter formulations in the mathematical framework. Chomsky (1995, p.1) explicitly states the main questions motivating the innovations of the Minimalist Program: “What are the general conditions that the human language faculty should be expected to satisfy?” and “To what extent is the language faculty determined by these conditions, without special structure that lies beyond them?”. These questions, of course, inspire more questions than answers resulting in concerns such as the place of the language faculty in general cognition and the conditions that the cognitive system imposes on the language faculty.

Chomsky (1995, p.3) reestablishes as the goals of linguistic theory to clearly define the structure of language, to specify how it yields utterances, and to account for how it is established in the mind/brain of budding speaker/hearers. This last goal, in particular, once again raises the issue of economy, an important factor in the progress of syntax in general. The Minimalist Program aims to establish a more economical framework for describing and explaining language derivation. A theory involving fewer rules or operations is preferred to a theory involving many rules and operations, just as simple rules and operations are preferred over complex ones. And a theory that derives structural descriptions using fewer operations and ones that are exclusively
obligatory, that only apply when absolutely necessary, is preferred over a theory involving a greater number of operations in derivation. Driven by economy, instead of utilizing intricate chains of specific transformations in phrase structure formation, the Minimalist Program considers more general transformations that can be applied more universally. Principles and Parameters did this with move α. The Minimalist Program builds off of that simplification and reduces the transformational component of the model even further.

Chomsky (1995) proposes that specific languages do not have rules governing construction of their clauses and phrases assuming the theory of Principles and Parameters is correct. Thus, a particular language is a set of conditional parameters in an unchanging system of principles of UG, and not a collection of rules. Instead of being a set of rules, language is a system of principles and language specific parameters. C-command and government seem to be universally present in the grammars of all languages, and are therefore likely part of the set of universal principles of syntax. The structural description of a language is initially the set of representations at the level of D-structure, representations at the level of PF, as well as representations at the level of LF, which all satisfy the conditions of the language.

However, Chomsky (1995, p. 219-220) rejects the ideas that D-structure and S-structure are present in the Principles and Parameters model. Instead of these levels each playing a vital role in language structure, a language can be described by the procedure that fabricates the representations (π, λ), where π is a representation at PF, and λ is a representation at LF. If an expression has a PF, LF and can be interpreted, Full Interpretation (FI) (which is a reinterpretation of the Projection Principle discussed in chapter 6) is satisfied. If FI is satisfied at
both interface levels, then an utterance is said to converge, and if not, it crashes. PF and LF do not interact, and lexical properties are not related to interface levels.

7.1 General Properties of the Minimalist Program Framework

Chomsky (1995, p. 225-226) initially redefines “syntax,” keeping within the limits of the economical requirements of the minimalist framework. A contrast is established between the lexicon, which houses information about lexical items, and the computational component. The computational component of human language (C\textsubscript{HL}) applies operations to form grammatical structural descriptions. C\textsubscript{HL} maps lexical choices to (\pi, \lambda). \pi and \lambda are derived from the same array of lexical choices, which indicates the lexical items chosen and how many times they are selected by C\textsubscript{HL} in the derivation.

A derivation is formed by C\textsubscript{HL} applying to a numeration\textsuperscript{31} (N) thus forming a sequence S of symbolic elements (\sigma_1, \sigma_2, …, \sigma_n) which halts when \sigma_n is a pair (\pi, \lambda), and S converges when it satisfies FI. The derivation progresses from the initial state by applying grammatical transformations to each sequential progression of the derivation. Once a set of well-formed constituents of LF and PH are established, the derivation ends, and we have a well-formed structure assuming FI is satisfied. C\textsubscript{HL} recursively forms \textbf{syntactic objects} by applying grammatical rules to formed syntactic objects and items in N. These syntactic objects are collections of lexical properties of their components (Chomsky 1995, p. 225-226).

\textsuperscript{30} In more recent work in the Minimalist Program, also how the lexicon interacts with the computational system. We will discuss these later in this chapter.

\textsuperscript{31} Numeration is simply the process of assigning a numbering order to a series of constituents (i.e. this one is first, this one is second, this one is third, and so on.)
7.1.1 Syntactic Objects

To form \( \lambda \) from \( N \), \( C_{HL} \) forms the set of syntactic objects \( \{SO_1, SO_2, \ldots, SO_n\} \) called \( \Sigma \). \( \Sigma \) can be interpreted at LF only if it consists of one syntactic object. Hence, there must be an operation that combines two already formed syntactic objects\(^{32} \). This operation is labeled \textbf{Merge}. Merge is the realization of \textit{move} \( \alpha \) in the Minimalist Program. Building on the simplification of the Principles and Parameters model, a more economical operation in conceived. Merge operates as many times as needed to ultimately form one final syntactic object containing all syntactic objects previously formed by the computational component, \( C_{HL} \). The resulting syntactic object must satisfy FI in order to converge (Chomsky 1995, p. 226).

A complex syntactic object \( K \) is formed from the set \( \{\alpha, \beta\} \) where \( \alpha \) and \( \beta \) are lexical components. Where \( \Sigma \) is a set of sequential syntactic objects where the terminal \( SO_n \) is used as the ultimate output of the derivation, \( K \) is any complex syntactic object that needs a label to indicate what type of a syntactic object it is. Because different parts of speech behave differently in structural descriptions, \( K \) must contain a label, \( \gamma \), indicating what type of category it is headed by. \( \alpha \) and \( \beta \) are eliminated as syntactic objects when Merge is applied, and they become constituents of the syntactic object \( K \). Thus, \( K \) consists of \( \{\gamma, \{\alpha, \beta\}\} \). Merge, which forms \( K \), is a recursive operation. The label of \( K \), \( \gamma \), is determined derivationally, and set when Merge is performed on \( \{\alpha, \beta\} \). \( \gamma \) is of the same type as either \( \alpha \) or \( \beta \), which ever one projects as the head of the syntactic object. Once the label \( K \) is established, \( K \) behaves as a syntactic object that is of the established label during the remaining computation. For example, if \( \alpha \) is a V and \( \beta \) is a N, \( K \) will be \( \{V, \{V, N\}\} \) (Chomsky 1995, p. 243, 246-247).

\(^{32} \)This is the reduction of X-bar theory to its minimalist counterpart.
7.1.2 Merge and Economy

**Merge** is an important operation in phrase structure formation and movement. Movement was previously achieved via move α after a lexical items were inserted in the phrase structure by another operation. But, in the Minimalist Program, movement utilized Merge, which is also used in phrase structure formation. The operation move α is eliminated and replaced by an operation that is used elsewhere in the derivation process, hence the model becomes more economical because there is one less operation to deal with. Merge combines individual syntactic objects to form a set of (complex) syntactic objects (Chomsky 1995, p. 226). For example, the syntactic objects *on* and *top* Merge together to form the set {*on*, *top*}.

Chomsky (1995, p. 226) asserts that the set formed by the application of Merge is a new syntactic object with syntactic object subparts and if a local grammatical relation cannot be formed, then economy prevents the formation of the set. Thus, sets with an internal grammatical connection such as {*at*, *home*} can be formed. The principle of economy requires that a set of lexical items have a local grammatical relation in order to form a syntactic object. This prevents sets with no local grammatical relation such as {*at*, *house*} to form. Phrase structures are formed using Merge, which welds two constituents together to form a larger constituent. Lexical items are taken from the lexicon and are either Merged with an already existing structure or start a new structure by merging with another lexical item. Merge is thus a recursive operation in that it applies to individual lexical items and larger structures that have themselves been formed by the Merge operation, as both are syntactic elements.

The Merge operation starts with two lexical items and converts them to a single phrase structure. Thus, Merge satisfies the definition of a mathematical transformation because when
applied to two constituents from the set of lexical items, it maps them to the set of phrase structures. Also, the same Merge operation is utilized to form more complex phrase structures when applied to two existing phrase structures; i.e Merge maps two elements from the set of phrase structures back into the set of phrase structures as a more complex structure. The recursive property of Merge establishes a valuable economy in that, not only is this single operation useful in a vast array of circumstances, but it is utilized over and over in the formation of a single phrase structure in the same spirit as its predecessor *move a*.

Because $\pi$ and $\lambda$ have different compositions, it is assumed that at some point in the derivation, the computation of the structural description diverges and LF and PF continue by applying their operations to their versions of the structural description, $\lambda$ and $\pi$ respectively. The point where $\pi$ and $\lambda$ diverge is spell-out, and after spell-out $\pi$ and $\lambda$ do not interact. Any operation can apply at any stage in the derivation of $\lambda$, but the derivation of $\pi$ does not have the same freedom. Spell-out occurs at some point in the construction of $\lambda$, and the covert component (formerly the logical component) continues to from $\lambda$ while the phonological component forms $\pi$. Before spell-out, the $\lambda$ is derived by the overt component. Spell-out removes all information not relevant to the formation of $\lambda$, forming $\Sigma_\lambda$ and the “covert component”(Chomsky 1995, p. 229) applies operations to $\Sigma_\lambda$ in the same fashion used in the formation of $\Sigma$ before spell-out. At the same time, the phonological component maps $\Sigma$ to $\pi$ where features now irrelevant to the derivation are removed. Select$^{33}$ is not available to the phonological component, but it is available to the covert component. However, the derivation will not converge at LF if an item with phonological features is selected. At some point in $\Sigma$, $\lambda$ and $\pi$ diverge to form separate

---

$^{33}$ Define select pg. 207-209
versions of the syntactic object being derived. Once this happens, all of these changes are made to \( \lambda \) and \( \pi \) (Chomsky 1995, p.225-229).

7.1.3 Phrase Structure

The emphasis of the Minimalist Program is minimalism and simplicity, hence X-bar theory developed in Chomsky (1970) is reduced to “bare phrase structure”. In this model, lexical items (maximal projections) and the heads projected from them have no distinction. There are no bar levels, and categories have the properties of the lexical items that determine them (Chomsky 1995, p. 249).

The phrase structure is generalized informally by the following structure.

```
where ZP = \{z, \{z,w\}\}, X’ = \{x, \{x,y\}\}, XP = \{x, \{ZP, X’\}\}. No status is given to the labels of the roots which is a deviation from the phrase markers used thus far. The only functioning elements are the nodes of the representation, namely x, y, z, w, X’, and XP (Chomsky 1995, p. 247).```
7.1.4 Copy

Copy is an operation used when extracting items from the lexicon. Clearly, when sets of syntactic objects are Merged together, the lexical items Merged are not removed from the lexicon. Rather, copies of the lexical items are Merged into the new set (Chomsky 2000, p. 114-115). The copy transformation maps lexical items from one set to another. For example, in the formation of sets of syntactic objects, the operation copy is applied to the set of syntactic objects in the lexicon and moved to N, the numeration set which will be used to form grammatical phrases, and each input item produces a unique output. Because copy maps selected constituents of one set onto a different set, it satisfies the definition of a mathematical transformation.

7.1.5 Movement

Move $\alpha$ is another operation involved in the derivation of structural descriptions. Given a structure $\Sigma$ with constituents $K$ and $\alpha$, $\Sigma'$ is formed by raising $\alpha$ to the target $K$ and replacing $K$ with $L$ which is composed of the set {$γ, \{α, K\}$} while leaving every other part of $\Sigma$ static. The movement operation introduces another $\alpha$ into the structural description via the operation copy which duplicates syntactic objects. The original $\alpha$ and the introduced $\alpha$ are distinguished by their position in the phrase structure (Chomsky 1995, p. 250-253).

Movement as a combination of the Merge and copy operations satisfies the definition of “transformation” because compositions of transformations are transformations. The application of the composite move operation moves phrase structures from one set to another. For example, the sentence above is moved from the set of phrase structures with nothing in the specifier
position of the TP to the set of phrase structures with a DP in the specifier position of the TP.

There is only one output phrase structure corresponding to each phrase structure that the
composite move operation is applied to, hence the move operation is one-to-one. Therefore, the
move operation as a composition of the copy and Merge operations is a transformation.

7.1.5.1 Delete

Delete α works along with move α by making a constituent α invisible to the interface and
leaving the rest of the structure unchanged (Chomsky 1995, p. 253). Deletion satisfies the
definition of a mathematical transformation in that it alters an input variable by removing its
visibility to the phonological component and making it a trace of the original constituent.
Deletion moves constituents with all lexical properties in tact to the set of trace elements,
satisfying the definition of a mathematical transformation.

7.1.5.2 Chains

Movement forms a chain (CH) describing the path the moved constituent took to arrive at
its destination. The chain formed by the movement of α in the example below is CH = (α, t(α)),
where t(α) is the trace of α (as we noted earlier in Principles and Parameters) (Chomsky 1995, p.
253).

Chomsky (1995) largely represents the initial state of the minimalist program. From
here, richer theories are developed as problems/issues emerge. New approaches branch off in
response to these initial ideas, one of which is distributed morphology, which seeks to explain
concisely how lexical items (LIs) are inserted with the most efficient/minimal method and details their status as syntactic objects.

7.2 Distributed Morphology

Halle and Marantz (1994, p. 275) describe the basic unit of morphology (or atom) as the Vocabulary item as contrasting with the lexical entry. The Vocabulary item contains semantic, syntactic, and morphological features, which identify which of the Vocabulary item’s phonological features will be inserted into a lexical head, as well as the item’s phonological features. This is different from the established Minimalist Program concept of Lexical Items, which are not necessarily a collection of features that determine insertion. The presence of Minimalist lexical items drive the derivation of syntactic objects in that they create labels and their features motivate movement. We will see below how the distributed morphology model contrasts.

There are three properties of Vocabulary items: (i) Late Insertion, (ii) Underspecification, and (iii) Syntactic Hierarchical Structure “all the way down” (to terminal nodes in the structure). The property of Late Insertion insures that the correct Vocabulary item is inserted into a given terminal node and prevents any modifications from being necessary to adjust the case/gender/etc after insertion. This is because the phonological features of Vocabulary items are only introduced after all syntax computation is completed. Vocabulary items do not contribute any syntactic or semantic information to a terminal node, they only add the phonological features of the specified Vocabulary item (Halle and Marantz 1994, p. 275).
Halle and Marantz (1994, p. 279) argue that Late Insertion is more economical than the early insertion advocated in other models of the operation of the lexicon where lexical entries provide all of the syntactic and semantic information needed by syntax and LF. In Distributed Morphology, no changes are made to constituents once they are in a structure, increasing economy.

Insertion of Vocabulary items satisfies the definition of a mathematical transformation because by inserting a Vocabulary item with matching features into a terminal node, a structure is essentially being moved from the set of grammatical structures without Vocabulary items to the set of grammatical structures with Vocabulary items. And the principle of Late Insertion ensures that this mathematical transformation is applied at most once per terminal node, again very mathematical and elegant.

The principle of Underspecification determines which Vocabulary item is inserted because it states that the set of features of the selected Vocabulary item must be a subset of the features of the terminal node. It is useful to note that if two sets are equivalent, then they are subsets of each other. Hence, if a Vocabulary item has the same set of features as a terminal node, then its features are still a subset of the set of features of the terminal node. More than one Vocabulary item can have a set of features that are a subset of the terminal node’s set of features. In this case, the Vocabulary item that is most highly specified is inserted into the terminal node (Halle and Marantz 1994, p. 276).

Underspecification is mathematical in the sense that the subset requirement is essentially borrowed from math. And the existence of a defined method of tie-breaking preserves the
adherence to the principles of mathematics. It even seems possible to program computers to perform Vocabulary Insertion as it is defined by Halle and Marantz.

The final principle, “syntactic hierarchical structure all the way down”, simply states that Vocabulary items are inserted into the terminal nodes of hierarchical structures that have been constructed by syntax. Halle and Marantz argue that this approach describes word formation purely syntactically (Halle and Marantz 1994, p. 276-277).

The Syntactic Hierarchical Structure principle (from the previous paragraph) specifies essential information about Distributed Morphology while not disrupting its conformity to the mathematics. Distributed Morphology is essentially a mathematical theory of word internal structure. With this new innovative way of looking at phrase structure derivation, the Minimalist Program look at syntax from alternative angles and respond with innovative approaches to derivation.

7.3 Minimalist Innovations

Chomsky (2000) establishes principles which describe the computational complexity of a realistic linguistic system. The complexity of the computational process is relevant to the viability of a cognitive system due to the resources available in the mind/brain. The computational system must explain all cases and be capable of representing all possible utterances in a quick fashion. The growth of computational complexity must be limited; it must not grow without bound. The choices and decisions made in the computation system should be subject only to UG.
Chomsky (2000) continues to describe phrase structure formation as involving the distinct operations Merge, as defined above, and agree, which “establishes a relation” between lexical items and features. Move combines Merge and agree, taking a constituent already present in the phrase structure and merging it with another position while also applying agree to satisfy feature checking. Because move is more complex, it does not comply with elegant theory design, and is highly restricted, applying only when Merge or agree cannot. This doesn’t seem to have any tangible consequences mathematically. Operations in mathematics are typically independent. There’s not really anything that would apply if another operation does not. But, its also not problematic in any way. This seems to be an economy issue, not one that raises the concerns that other sorts of constraints raised.

However, the Minimalist Program does encounter problems with adherence to the laws of mathematical transformations. Chomsky (2000) introduces the idea of a grammar allowing structures to be created which will eventually violate some grammar rule and crash within their formation. Grammatical structures will converge to a grammatical phrase structure if they converge on all interface levels.

Structures converge on a given interface level if they contain all of the necessary and sufficient information required on that level. Convergence assumes that a structure contains all of the constituents and the form required at a given interface level without the inclusion of extraneous information. If a structure does not satisfy the requirements of convergence, then it crashes at a given interface level (Chomsky 2000). In Chomsky (1995), this is equivalent to Full Interpretation not being satisfied. This aspect of the model is both mathematically and economically unsatisfying. Ideally, all produced structures should converge and the
computational system should not produce anything that will eventually crash. And although there are plenty of examples of utterances which are clearly ungrammatical after production, they are a result of memory limitations and mid-structure production mind changes, and need not be addressed by a theory of the computational system.

This property also violates the principles of mathematical transformations in that mathematical transformations exclusively produce constituents that belong in a set and violate no rules of inclusion. For a transformation to produce an output that does not ultimately conform to the set it is mapped to is a violation of mathematical principles.

The production of structures that are not well-formed is wasteful and unlikely to mirror the operations taking place in the mind/brain. (Frampton and Gutmann 2000) argue that in order to form a more optimal derivational system, all crashing after or during the progression phrase structure derivation must be eliminated. They propose that phrase structure which will ultimately lead to ungrammatical phrase structures should be eliminated in the first steps of derivation.

The most efficient computational system is one that will not ever allow phrase structures that will become ungrammatical to form even in the initial steps of derivation, as described by the Strong Minimalist Thesis explored in the next section. The computational system is driven by necessity and hence does not produce anything ungrammatical. Lexical items and phrase structures are only inserted in an existing phrase structure by necessity.

Driven by pure necessity in language, Chomsky introduces a new motivation for language (the Strong Minimalist Thesis), and modifies the theory of syntax to contain only necessary information.
7.4 The Strong Minimalist Thesis

Chomsky (2008, p.135) considers what he coins the Strong Minimalist Thesis (SMT), which states that given the requirements of the Faculty of Language, language is the optimal solution subject to these requirements. The SMT is not expected to hold fully, but is a tool to formulate the theory of language. In particular, when facing a deviation from the SMT, the fact that the principle in question departs from the SMT tells us that it needs to be looked at carefully with extra scrutiny to determine if the principle is in fact essential.

With the SMT as a goal, the theory of language is formulated to be as minimal as possible. The interface levels available for the Sensory-Motor(SM) and the Conceptual-Intentional(CI) systems are essential for the process of phase formation. These systems are independent of language, and their properties do not contribute to the theory of language (Chomsky 2008, p. 137).

Chomsky (2008, p. 138) eliminates as not essential, and therefore not needed, properties formerly considered sound: traces, indices, and other descriptive tools, reducing the theory to only what is indispensable. Merge is considered in the simplest imaginable form: Merge takes two Syntactic Objects(SO’s) and combines them, and leaves the SO’s unchanged. Merge cannot add or delete any features of the SO’s or break them up in any way. This definition of Merge continues to satisfy the definition of a mathematical transformation.

Merge can only be applied to a Lexical Item(LI) if it has a property, called a feature, that motivates the application of the transformation. The feature justifying the application of Merge is called an Edge Feature(EF). When an LI is Merged with an SO, the set \{LI, SO\} is formed and SO is the LI’s complement. Merge now defines the entire computational system,
accommodating the infinite and recursive nature of language, and eliminating the need for d-
structure and s-structure (Chomsky 2008, p. 139).

There are now two types of Merge. The first is External Merge (EM), which is utilized when
neither of the two objects is a part of the other. EM is used in insertion. The alternative is
Internal Merge (IM), used when one of the Syntactic Objects is embedded in the other. For
example, if X is embedded in Y, and X is Merged with Y, IM is used and the result is $Z = \{X,Y\}$
with two copies of X in the set. IM is the new Move, and it creates copies of the SO being
moved (Chomsky 2008, p. 140).

Both Internal and External Merge satisfy the definition of a mathematical transformation.
They continue to operate constituents in the set of Syntactic Objects resulting in new constituents
of the set of Syntactic Objects.

The set of syntactic relations is reduced to two: set membership (based on Merge), and
probe-goal. Probe-goal is Agree or IM, when structure-internal, and EM, when structure-
external. Each SO has a label which contains all information relevant to computations. For EM,
the label selects/is selected, and for the internal operations, the probe seeks a goal (such as
satisfying Agree) (Chomsky 2008, p. 141). So, each operation is minimal. There is no
“combination” of operations as there was even in the mathematically satisfactory transformations
of Aspects (Chomsky 1965, discussed in chapter 3).

Set membership clearly conforms to the principles of mathematics in that set membership
is a mathematical concept, and is being used true to the mathematical definition here. Probe-goal
is also mathematical because it is a systematic way of motivating the application of Merge. It
doesn’t necessarily change the structures in any way, but establishes how/why Merge applies.
“Minimal search” identifies labels during computation. There are two operations which work in parallel to determine an SO’s label. If H is a head, H is the label in \{H, a\}. And b is the label in \{a, b\} if a is Internally Merged to b (Chomsky 2008, p. 145).

**Transfer** operations map constructed SO’s to the phonological component/SM interface, Spell-Out, as well as the semantic component/CI interface. At this point, SO’s are now called phases, and the phases are the same for both transfer operations. Transfer deletes any information that is irrelevant to the interface the phase is being mapped to. This means that, when an SO is mapped to the semantic component, any information that is not needed by the semantic component will be deleted. Likewise for mappings to the phonological component (Chomsky 2008, p. 142-143). Transfer satisfies the definition of a mathematical operation because it maps objects from the set of SO’s to the set of phases in either the phonological or semantic components. and deleting irrelevant information in the process is perfectly acceptable for mathematical transformations.

In the context of the goals set by SMT, Chomsky continues to chip away at the superfluous components of the Minimalist Program and more innovative reductions are made in Chomsky (2013).

Chomsky (2013, p. 36-37) argues that order is not essential to syntax, and like phonology, is only applicable at the Sensory Motor(SM) interface where it is needed for communication. Language is foremost used for thought, where according to his argument, order is potentially dispensable. Not much is known about the structure of the brain in reference to how its structure effects language and language learning. Hence, when two alternative ideas are being compared, there is a significant advantage to the simpler, less computationally complex proposal. Examples
of ambiguity, such as *The Chickens are too hot to eat*, show that language is not optimally designed in that language semantic/pragmatic interpretation is not always the priority. In fact, these perception difficulties tell us that computational simplicity is valued over communication (Chomsky 2013, p. 41).

Projection/labeling is not supported by empirical evidence of its essentialness to syntax. It has been passed down from traditional structuralist grammar as a useful tool in the exploration of language and is not necessarily a vital part of the theory of syntax (Chomsky 2013, p. 37).

Abandoning the traditional projection approach does not eliminate the need for many established components of the theory, but does, in some cases, change the way these components are defined. Merge takes constructed objects X and Y, and combines them resulting in the new object \( Z = \{X,Y\} \). X and Y are not ordered in Z and are not changed by the application of Merge. And IM and EM are still considered essential (Chomsky 2013, p. 40).

The quintessential innovation of Chomsky (2013) is the elimination of order, which is in line with the principles of mathematics. In mathematics, sets are unordered, so eliminating the order requirement of a SO, which is now simply a set, moves the theory closer to math.

The Minimalist Program set as its goal finding the most realistic theory of syntax, and the latest innovations accomplish that by eliminating as much superfluous computation/information in order to accommodate the limitations speaker/hearers face. The Minimalist Program moved in and out of mathematics, but the current theory conforms to the principles of mathematics.
Chapter 8: Conclusion

Chomsky (1881, p.2) stresses the abstract mathematical form of syntax by stating “the abstract may relate to transformational grammar rather in the way that modern algebra relates to the number system.” In abstract algebra, structures are studied rather than any particular number system. And although, the number system does correspond to some of the particular structures, numbers are typically used more as an illustration of the structures to which they correspond. Likewise, syntax is a study of the structures of language and language development. In order for such a complex system to be so easily used by such a vast group, there must be a strong inherent structure, and this is what syntax strives to uncover. Attempting to get closer to this structure, Chomsky (1955) was influenced by mathematical concepts, and because it was such an innovation in linguistic theory, I have traced the influence and relevance of mathematics in syntax from the big leap of Generative Grammar in the 1950’s through the most recent definition of the research program, the Minimalist Program.

Mathematical transformations are the most relevant math concept for studying the coherence of the progression of the models of syntax. In Chapter 2, we saw that the model developed in Chomsky (1955) was consistent with mathematical transformations, and the grammar posited acted as a system of mathematical transformations mapping constituents from one set to another. This seminal development in linguistic theory is the basis for this thesis. Incorporating mathematical transformations into syntax lends insight to the underlying structures of utterances and the development of the language system by language learners. Because a
model consistent with the principles of mathematics is an indication of a well-formed model, the relevance of mathematics was evaluated in subsequent works of syntax.

In Chapter 3, the further development of a mathematical model of syntax was traced. The Chomsky (1965) model was the beginning of the incorporation of semantics into the model. A distinction between the semantic component and the syntactic component was developed in deep structure and surface structure. This innovation was consistent with the principles of mathematics, but raised questions about the relevance of semantics to syntax that lead to a deviation from a mathematical model.

In the 1960’s and 1970’s semantics moved to the forefront of the study of syntax, the topic of Chapter 4. The issues of overgeneration and semantic interpretation were the focus of this period because these were the deficiencies of the existing theory in modeling speaker intuition. A constraint based model (Ross 1967) was developed in which the domains of grammatical transformations were restricted, consistent with the principles of mathematics. Ross (1967) also proposed filters, which would eliminate utterances after they are formed based on speaker acceptability judgements, and such filters violate the principles of mathematics. Scalar acceptability of utterances was also posited in Ross (1967). Scalar judgments allow the interpretation that some utterances are well-formed, less well-formed and not well-formed and this is not mathematical. As discussed, in mathematics, a constituent either has or does not have a property; there is no scalar acceptability of mathematical concepts (i.e. a number is either prime or composite; there are not numbers that are a little bit prime). Moving away from consistency with mathematics makes syntax less desirable from an empirical perspective given the memory and computational limits of the mind/brain where language resides.
During this period, Katz and Postal (1964) proposed that all semantic information is present at deep structure, and Katz and Fodor (1963) proposed that semantics determines structure. These propositions paved the way for generative semantics. Generative semantics held meaning as integral to the final structure of utterances and speaker judgments. As a result, case theory was born (Fillmore 1968). These developments mark a clear deviation from the structure based approach advocated by Chomsky from 1955. At this point, mathematics is on the back burner of thought relating to the model of syntax. This allowed for the exploration of some core issues of syntax and was integral to finding solutions to the problems exposed.

In chapter 5, the response to generative semantics by those who advocated a structure based model was considered. Chomsky (1970) noted that many of the proposed semantic issues proposed by generative semantics can be explained by examining structure in mathematical way. Chomsky (1973) advocated performance filters which restrict the structures to which grammatical transformations can apply so ill-formed utterances are not produced. Both proposals (Chomsky 1970 and 1973) conform to mathematics.

Jackendoff (1972) proposed wellformedness conditions which mark as unacceptable utterances that are structurally well-formed, but don’t have a sensible meaning. This approach deviates from mathematics by eliminating structures after their formation; this is similar to the proposal of Ross (1967). And Lakoff (1971) offered the solution that individuals have different experiences, knowledge, and opinions of the world and those differences inspire an individual’s grammaticality judgements. Graded speaker judgments, where language users rate utterances as more or less well-formed much like Ross’s (1967) idea, are incorporated by Lakoff (1971). Once again, such ideas violate the principles of mathematics. Chomsky and Lasnik’s (1977)
optimality (where constraint violations are ranked and the least unacceptable violations are considered well-formed) and surface filters (ill-formed structures are eliminated after they are formed) are also not mathematical. Thus, reducing the relevance of mathematics to the model and the clarity and economy that mathematics brings.

In Chapter 6, the priorities of the original model of generative grammar are brought back to the forefront of syntax while addressing the issues raised by generative semantics in the Principles and Parameters model. A subset of language parameters from UG are selected by a child as he or she develops language based on the primary linguistic data (PLD) to which he or she is exposed (Chomsky 1981). This model conforms to the principles of mathematics in that restricting the language faculty to contain language specific information is analogous to restricting the constituents in a set (i.e restricting a set to contain only even numbers). This change can be thought of as a restricting mathematical transformation mapping linguistic knowledge from a more general set onto a language specific set by eliminating unacceptable parameters. The transformational component is reduced to the single transformation move $\alpha$, driven by case, with a restricted domain in Chomsky (1980) and grammatical transformations are once again mathematical transformations mapping constituents among sets.

The Minimalist Program, which builds on the foundation of Principles and Parameters and was the focus of chapter 7. Merge, a recursive operation which combines syntactic objects to form a set, replaced move $\alpha$ (Chomsky 1995). Movement is reduced to copy and Merge, both of which as well as their compositions (back to back use) satisfy the definition of a mathematical transformation. Distributed Morphology (Halle and Marantz 1994) is a fascinating mathematical
model of word internal structure exemplifying the elegance and beauty of mathematics in syntax and the clarity mathematics brings.

However, the Minimalist Program moves away from its mathematical foundation by allowing phrase structures to crash during formation if all requirements for convergence do not exist in Chomsky (2000). A model that allows the formation of structures that will be deemed ill-formed during their derivation is inherently not economical, and a model that does not allow for the formation of extraneous structures is more economical and more likely to mirror the mind/brain (as we’ve seen in the discussions of overgeneration). At the same time, Chomsky (2000) emphasizes the importance of limiting the complexity of the computational component, a goal mathematics will help achieve.

Chomsky (2008) develops as the goal of syntax the Strong Minimalist Thesis (SMT) stating that only what is absolutely indispensable should be included in the model, eliminating projection rules. Chomsky (2013) notes that ambiguity (*The chickens are too hot to eat*) points to simplicity being valued over clarity in the mind/brain. He posits that order is not essential, and syntactic objects (SO’s) are unordered sets, which is consistent with mathematics where sets are inherently unordered. The operations Transfer (which maps structures to the phonological component/SM-interface, Spell-Out, and the semantic component/CI-interface) and Merge satisfy the definition of mathematical transformations. Seen as a collection of unordered sets related by mathematical transformations, this model of syntax is the closest to mathematics since and including the model at the inception of generative grammar. Hence, the Minimalist Program returns its mathematical underpinnings on which Generative Grammar was founded.
At times, each progression has lost sight of the mathematical underpinnings of Generative Grammar in order to uncover and address problems of the theory of syntax, which of course is the goal of scientific discovery. It is through the continuous formulation of new questions and observations of shortcomings of accepted thought as well as challenges posed by the new that progress in understanding is made. In other words, “if we are satisfied that an apple falls to the ground because it is its natural place, there will be no serious science of mechanics” (Chomsky 1995, p. 4).
References


Frampton J, and Gutmann, S. 2000. Agreement is Feature Sharing.


Appendix: Definitions

each definition first appears on the page number listed in parentheses

**composition**: (26) A sequence of transformations applied in a given order.

**computational component**: (81) applies operations to form grammatical structural descriptions.

**concatenation**: (24) elements being joined together to form a single new element. For example, *un* and *happy* are concatenated (illustrated with the symbol $\cap$) at the word level to form the constituent *unhappy* ($un \cap happy$).

**converge**: (81) a phrase structure converges when Full Interpretation is satisfied.

**copy**: (86) an operation used when extracting items from the lexicon.

**crash**: (81) a phrase structure crashes when Full Interpretation is not satisfied.

**deductive**: (5) assuming several premises to be true and arriving at conclusions based on the logical consequences of the assumed premises.

**deep structure**: (6) a structure distinct from what is superficially accessible in a single string, which governs semantic interpretation.

**descriptive adequacy**: (2) describing the idiosyncrasies and commonalities of language.

**domain**: (12) the set on which a mathematical transformation can be applied.

**economy**: (5) the simpler the descriptive mechanism, the more it is likely to describe and plausibly, thus, it’s more falsifiable.

**explanatory adequacy**: (3) describing how individual speaker/hearers come to know how to use/understand language

**filters**: (47) a formal grammar rule of languages which places acceptability judgments on the phrases produced by grammatical transformations and rules.

**Full Interpretation (FI)**: (80) satisfied if an expression has a PF, LF and can be interpreted.

**generative grammar**: (4) a set of rules that provides structural descriptions for utterances. The core claim is that every speaker of every language has mastered a generative grammar in order to produce and understand utterances, and to distinguish between those that are part of the language and those that are not.
grammar: (4) the set of rules used in the derivation of the elements of the set of all grammatical sequences.

grammatical transformation: (1) an operation that modified a grammatical constituent and maps it to a new set of grammatical elements.

ideal speaker/hearer: (15) conceived to have a complete understanding of a language in a homogenous speech community and be (in this ideal state) unburdened by such things as memory capacity, attention span limitations, and “erroneous” usage (i.e. mistakes by speakers such as false starts and incorrect usage). The ideal speaker/hearer as a representation of a speech community does not exist.

I-language: (19) an individual’s internalized knowledge of a language.

Inductive: (4) looking at an array of data and developing a theory about general behavior based only on the restricted array.

kernel: (29) the set of deep structures.

lexical items: (2) constituents of the lexicon, including words, morphemes, and idiomatic phrases.

lexicon: (6) made up of constituents with a system of lexical features, called lexical entries.

numeration: (82) the process of assigning a numbering order to a series of constituents (i.e. this one is first, this one is second, this one is third, and so on.)

mathematical transformation: (1) maps members of one set to another set. For example, the function f(x)=2x, where x is an element of the set of integers, is a mathematical transformation that maps members of the set of integers to the set of even integers.

Merge: (82) combines individual syntactic objects to form a set of (complex) syntactic objects.

minimal pair: (30) two linguistic structures apparently identical in all respects except one.

optimality: (68) ranks the violations of grammaticality allowing the least unacceptable structure to remain grammatical.

overgeneration: (32) producing structures that are not actually a part of a grammar.

parameters: (71) a set in a child’s developing - and thus language specific - language faculty. As a child develops language, he or she eliminates/includes language specific parameters based on
linguistic experience, namely exposure to the linguistic outputs of more skilled language users, such as parents.

**Poverty of the Stimulus (POS):** (4) However rich a child’s exposure to a language is, it will never include all the well-formed utterances that a child will be able to produce and understand of the course of his or her lifetime.

**Primary Linguistic Data (PLD):** (4) The language use that children are exposed to, and as a result, develop language.

**recursion:** (5) using a limited number of simple tools over and over again rather than many more complicated tools

**set:** (2) A collection of objects.

**set theory:** (1) A branch of mathematics studying sets and relationships among sets.

**Strong Minimalist Thesis (SMT):** (93) given the requirements of the Faculty of Language, language is the optimal solution subject to its requirements.

**surface structure:** (1 Ch3) regulate phonetic interpretation, and are closer to the more superficially accessible strings.

**syntax** (1) the examination of the theoretical structure of human language.

**syntactic objects:** (81) are collections of lexical properties of their components (i.e. phrases, lexical items)

**Transfer:** (95) operations map constructed SO’s to the phonological component/SM interface, Spell-Out, as well as the semantic component/CI interface.

**Universal Grammar(UG):** (14) an innate mechanism designed for language learning which includes innate information including a restriction on the form of possible languages.

**unmarked:** (68) a construction is unmarked when it occurs more typically than a more unusual alternative. Such as the present tense vs. the gerundive. Markedness has long been a problematic category marker.

**X-bar theory:** (60) a system of visually representing the underlying construction of phrases in which the categorial component adheres to construction rules.