Chapter 2: Brief Overview of the History of Generative Syntax[*]

Howard Lasnik and Terje Lohndal
University of Maryland

1. Background and outline

Scientific grammar goes back to the Middle Ages, and specifically the study, by Modistic philosophers, of language as a phenomenon independent of thought. In a sense, the tradition is even older, dating back to Classical Antiquity, and spanning several cultures – after all, every traditional writing system presupposes some serious linguistic theorizing. In the humanistic Renaissance, philosophers started worrying, also, about the relation between language and mind, and as the Ages of Exploration and Reason came to be, about the problem of creativity and what it reveals about the natural world – where according to Descartes it effectively constituted a "second substance". By the time Darwin began to revolutionize our thinking about human nature, philology was a profession in its own right, so much so that the discovery of the Indo-European ancestry and how it gave rise to hundreds of different languages served as a central inspiration in Darwin's evolutionary theory. Many of the theoretical insights of linguistics in the twentieth century date back to this modern tradition, particularly as coupled together with late nineteenth and early twentieth century developments in mathematical logic and philosophy more generally.

Saussure (1916) initiated contemporary structural linguistics, by emphasizing how language should be conceived as separate from what it is used for, and by concentrating on how language is, not how it changes. Bloomfield (1933), Wells (1947) and Harris (1951) developed structuralism further and Noam Chomsky's work developed in particular in immediate reaction to Harris's program. A fundamental difference between structuralism and generative grammar stems from the fact that Chomsky focused on those aspects of structure that make the system recursive, whereas structuralism left those for the realm of what we nowadays call performance. Structuralism in fact focused on finite levels of language, such as morphophonemics, where notions like "linguistic feature" or the paradigmatic inventory underlying phonemics came to be understood (see again especially Harris (1951)). But it was the syntax put to the side at the time that especially interested Chomsky, particularly since it was taken to address a key element in the problem of linguistic creativity. For this purpose, Chomsky borrowed from the axiomatic-deductive method in mathematical logic, developed a generation earlier in its computational formulation – more concretely via Davis 1958 (which had circulated as a draft much prior to its publication date). Chomsky systematized and generalized Emil Post's version of "recursive function theory" (see Post 1944), and eventually came to propose formal devices of his own ("transformations", see below).

Aside from these theoretical considerations pertaining to the precise structure of language and its implications, generative grammar from Chomsky's perspective always had a conceptual angle that informs the enterprise to this day: Syntax is seen as a natural system, somehow rooted in human psychology and biology. This point of view

[*] We are grateful to Juan Uriagereka for extensive help on an earlier draft and to Marcel den Dikken for his patience and encouragement, and to Marcel and an anonymous reviewer for very helpful comments that led to substantial improvements in the presentation.

constituted the bulk of Chomsky's reaction to behaviorism, his later exploration of complex forms of biology, and more generally his insistence over six decades on approaching linguistic structure with the same sorts of tools and attitudes that one should assume for an intricate biological phenomenon, like adaptive immunity.

All of Chomsky's work has centered on two fundamental questions:

(1) What is the correct characterization of someone who speaks a language? What kind of capacity is "knowledge of language"?

(2) How does this capacity arise in the individual? What aspects of it are acquired by exposure to relevant information ("learned"), and what aspects are present in advance of any experience ("wired in")?

Chomsky's earliest work, in the 1950's, raised and focused on question (1), since explicit and comprehensive answers to that question had never been provided before. Chomsky's answer posited a computational system in the human mind that provides statements of the basic phrase structure patterns of languages (phrase structure rules) and more complex operations for manipulating these basic phrase structures (transformations). This framework, and its direct descendants, fall under the general title Transformational Generative Grammar ("generative" meaning explicit, in the sense of mathematics).

In the 1960's, the research began to shift more towards question (2), and Chomsky called the theory that was developed the Standard Theory. Chomsky coined the term explanatory adequacy for putative answers to that question. A theory of language, regarded as one component of a theory of the human mind, must make available grammars for all possible human languages. To attain explanatory adequacy, the theory must in addition show how the learner selects the correct grammar from among all the available ones, based on restricted data. The theories of the 1950's and early 1960's made an infinite number of grammars available, so the explanatory problem was severe.

Through the late 1960's and 1970's, to enhance explanatory adequacy, theorists proposed more and more constraints on the notion 'possible human grammar'. Ross (1967) was a particularly influential and pioneering study looking at locality restrictions (see chapter 18). These moves were explicitly motivated by considerations of explanatory adequacy, though general considerations of simplicity also played a role. This culminated in the Principles and Parameters framework (chapter 4), and more specifically in the Government and Binding approach that Chomsky (1981) proposes. The latter led to a wide range of cross-linguistic research since a core part of the program involved comparative syntax and used comparative data to help refine theoretical definitions of terms like "government" and "binding".

At the same time as these developments took place, a number of researchers departed from Chomsky's specific approach. Generative semantics, in particular, was a very prominent theory in the late 1960's; today some generative semantics ideas have returned, as we will discuss below. In the early 1980s, non-transformational theories such as Lexical-Functional Grammar (Kaplan and Bresnan 1982; see chapter 6), Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum and Sag 1985; see chapter 7) and Tree Adjoining Grammar (Joshi, Levy and Takahashi 1975, Joshi 1985; see chapter 8) were also developed. We will say a bit more about these below and contextualize them to make it clear why they emerged and what the main differences are between these theories and the more mainstream Chomskyan theories.

In the late 1980s, Chomsky started to explore what has become known as the Minimalist Program, with its emphasis on simplicity in theorizing and on moving beyond explanatory adequacy in the sense of asking why the language faculty has the properties it does. This approach is most explicitly outlined in Chomsky (1995b). Recent and ongoing work by Chomsky (2000, 2001, 2004, 2007, 2008) and many others continues to develop this framework.

This chapter is organized as follows. Section 2 discusses the earliest generative approaches, namely those explicated in *Syntactic Structures* (1957) and *Aspects of the Theory of Syntax* (1965). We examine some relevant differences between these two theories and we discuss some general properties of transformations. Section 3 discusses the syntax/semantics interface in early generative grammar and beyond whereas section 4 is an overview of how phrase structure has developed from the early days of generative grammar until today. In section 5, we discuss the role in the evolving theories of rules and filters versus principles. Section 6 is concerned with derivations and the derivation versus representation issue. In Principles and Parameters theory, Chomsky explicitly introduced economy principles for the first time, and we give a summary of some of these in section 7. A few concluding remarks are provided in section 8.

2. The earliest generative approaches: *Syntactic Structures* and *Aspects*

Chomsky's earliest work developed in reaction to the structuralist work mentioned in the introduction. As a student of Zellig Harris, Chomsky was very familiar with Harris's program and he developed his own work in reaction to Harris (1951). Harris had one sentence transform into another. This approach was therefore not able to give any systematic explanation for the more abstract kind of phenomena Chomsky started to deal with in *The Logical Structure of Linguistic Theory* (*LSLT,* 1955) and *Syntactic Structures*. In order to deal with these phenomena, it is necessary to relate abstract structures to abstract structures. Let us now look at some of the characteristics of Chomsky's earliest work.

Infinity and structure are the fundamental characteristics of human language, and they can both be captured, in part, by way of a context-free phrase structure (PS) grammar. One such device (a Σ, F grammar in Chomsky's terminology) consists of:

(3) a. A designated initial symbol (or a set thereof) (Σ);
 b. Rewrite rules (F), which consist of a single symbol on the left, followed by an arrow, followed by at least one symbol.

A derivation consists of a series of lines such that the first line is one of the designated initial symbols, and to proceed from one line to the next we replace one symbol by the sequence of symbols it can be rewritten as, until there are no more symbols that can be rewritten. For instance given:

(4) a. Designated initial symbol (Σ): S

b. Rewrite Rules (F):

S → NP VP
NP → N
VP → V
N → John
V → laughs

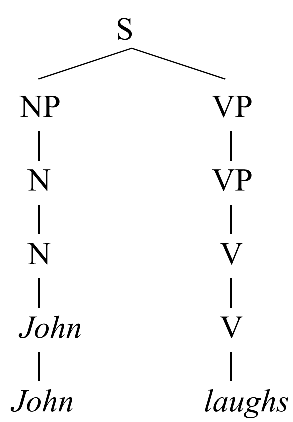
We can obtain a derivation as in (5):

(5) Line 1: S
Line 2: NP VP
Line 3: N VP
Line 4: N V
Line 5: John V
Line 6: John laughs

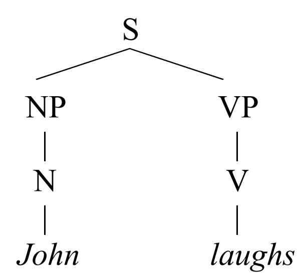
Chomsky (1965) called rules like the last two in (4), which rewrite a particular non-terminal symbol as a single terminal symbol, lexical insertion rules – a distinction not made in the theories of Chomsky (1955, 1957).

PS grammars capture constituent structure by introducing non-terminal (unpronounced) symbols. Given (5), we can connect each symbol with the symbol(s) it rewrites as. In this way we can trace back units of structure. After joining the symbols we can represent the derivation in the standard form of a tree as in (6a). Getting rid of symbols that are mere repetitions, we end up with the collapsed tree in (6b):

(6) a.



b.



More technically, a phrase marker for a terminal string is the set of all strings occurring in any of the equivalent derivations of that string, where two PS derivations are equivalent if and only if they involve the same rules the same number of times (not necessarily in the same order). This is a result that Chomsky (1955) proved by showing that for two PS derivations to be equivalent, they have to collapse down to the same PS tree. See section 4.1 for further discussion.

2.1. Transformations and generalized transformations

Finite-state machines can easily capture infinity, one of the two fundamental characteristics of human language (see Lasnik 2000 for much discussion), and if we move one level up on the Chomsky Hierarchy (Chomsky 1956), we can avail ourselves of PS grammars. These grammars are more powerful devices that capture both infinity and structure.

Interestingly, the theory in both *Syntactic Structures* and *The Logical Structure of Linguistic Theory* (Chomsky 1955, henceforth *LSLT*) did not have recursion in the base, that is, PS rules, or sequences of them, that allow self embedding. Instead, complicated structures, hence infinity, were created by special operations, called *generalized transformations*, which put together the simple structures generated by the PS rules. For example, to derive *John knew that Mary understood the theory*, first the separate structures underlying *John knew it* and *Mary understood the theory* were generated by the method described above; then a generalized transformation inserted the second of these structures into the first. Metaphorically, a generalized transformation grafts one tree onto another. Put differently, in this theory recursion was in the "transformational component".[1] In more recent times, Tree Adjoining Grammar (TAG) developed this approach further (Joshi, Levy and Takahashi 1975, Joshi 1985; see chapter 8) by arguing for a system of tree rewriting. In this theory, a derivation works on a set of predefined pieces of tree structure. These pieces are called elementary trees and they are expanded and combined with one another so that structures are built through generalized transformations. Still more recently, Frank (2002) suggested a way to integrate the minimalist approach to grammar suggested by Chomsky with TAG.

The structures created by phrase structure rules and generalized transformations could be altered by singulary transformations.[2] Singulary transformations apply to single P-markers and derived P-markers, which is to say that they apply to one tree. Chomsky showed how singulary transformations can explain the relatedness between, for example, statements and corresponding questions:

(7) a. Susan will solve the problem. → Will Susan solve the problem?
  b. John is visiting Rome. → Is John visiting Rome?

The members of each pair come from the same initial P-marker, with singulary transformations producing the divergent surface shapes. One of the great triumphs of the analysis of such pairs in *LSLT* is that it was able to use the same singulary transformation for the interrogative sentences in (7) and the superficially very different one in (8).

(8) Susan solved the problem. → Did Susan solve the problem?

---

[1] In languages like Hungarian, sentential complementation is typically "mediated" by a pronoun, as shown in (i).

(i) Janos azt tudja, hogy S
  Janos it.ACC knows that S

This property may provide retroactive support for the *LSLT* way of generating sentential complementation. Thanks to Marcel den Dikken (p.c.) for pointing out this fact to us.

[2] It should be noted that singulary transformations and generalized transformations could be interspersed. There were no constraints on when either could apply.

This was a significant achievement since the relations are felt by native speakers to be parallel, an otherwise mysterious fact. Chomsky also showed how in numerous situations, even properties of individual sentences cannot be adequately characterized without recourse to the descriptive power of singulary transformations. One major example involved the sequences of English auxiliary verbs, and the inflectional suffixes associated with them. The revolutionary insight here (and also in the analysis of (7)-(8)) was that these bound morphemes, especially the one carrying tense and agreement, are autonomous items as far as the syntax is concerned, capable of undergoing syntactic operations independently until eventually uniting with a verbal element (a process that came to be called Affix Hopping). The Affix Hopping transformation rises above the limitations of phrase structure (which at best can simply list the possible sequences) and simultaneously captures the generalizations about linear ordering of the elements, their morphological dependencies, the location of finite tense, the form of inversion and sentence negation, and the distribution of auxiliary *do*.[3] There was, thus, considerable motivation for this new device relating more abstract underlying structures to more superficial surface representations. In fact, one of the major conceptual innovations in the entire theory is the proposal that a sentence has not just one structure, closely related to the way it is pronounced, but an additional abstract structure (potentially very different from the superficial one), and intermediate structures between these two. This is fundamental to all the analyses in the Chomskyan system.

The organization of the syntactic portion of the grammar is as follows: Application of the phrase structure rules creates a P-marker, or, in the case of a complex sentence, a set of P-markers. Then successive application of transformations (singulary and generalized) creates successive phrase structure representations (derived P-markers), culminating in a final surface representation. The syntactic levels in this theory are that of phrase structure and that of transformations, the latter giving a history of the transformational derivation (the successive transformational steps creating and affecting the structure). The representations at these levels are the P-marker and the T-marker respectively. The final derived P-marker is the input to phonological interpretation, and the T-marker is the input to semantic interpretation.[4]

Let us consider some of the formal properties of transformations as they are stated in *Syntactic Structures*. Each transformation has a structural analysis (SA) and a structural change (SC). The SA characterizes the class of structures to which the transformation applies. The SC specifies the alterations that the process carries out.

A SA is a sequence of *terms* or a set of sequences of terms. Elements that can constitute a term are listed in a general fashion in (9).

(9) a. any sequence of symbols (terminals, nonterminals, and variables) *or*
b. a set of sequences of symbols *or*
c. a Boolean combination of these

SCs are able to carry out the following elementary operations:

---

[3] For extensive discussion of this analysis, see Lasnik (2000, chapter 2).
[4] We need the T-marker as interface with semantics because the final derived P-marker typically lacks information relevant to meaning, for example grammatical relations if Passive has applied.

(10) a. adjunction of one term to another (to the right or the left)
b. deletion of a term or sequence of terms
c. adjunction of new material that was not in the structure before to a term
d. permutation

A SC for Chomsky was a set of elementary operations.

Other properties of transformations are that they are ordered, and that they are specified as being optional or obligatory. For some transformations it is crucial that we be allowed but not required to apply them; for others it is necessary that we be required to apply them. Lastly, the transformations in *Syntactic Structures* also occasionally had a global dependency: They can refer back to any other stage of a derivation.

We will not go through an example of an early generative syntactic analysis here but instead refer to Lasnik (2000: 53ff.) for a thorough illustration of several early transformations.

2.2. Chomsky (1965)

Chomsky (1965), henceforth *Aspects*, presented a revised conception of the grammar, based on an alternative way of constructing complex sentences, one that Chomsky argued was an advance in terms of simplicity and explanatory adequacy over the one in *LSLT*. In the *LSLT* framework, as discussed above, the phrase structure rules produce simple monoclausal structures, which can then be merged together by generalized transformations. Generalized transformations were thus the recursive component of the grammar, the one responsible for the infinitude of language. In the alternative view, the phrase structure rule component itself has a recursive character. Consider the complex sentences in (11).

(11) a. Mary reads books.
b. John thinks that Mary reads books.
c. Susan said John thinks Mary reads books.

By adding a recursive "loop" to a standard set of phrase structure rules, we can directly create the possibility of ever longer sentences. Such a rule is given in (12).

(12) VP $\rightarrow$ V S

Under this approach to sentence embedding, unlike that in *LSLT*, there is one unified structure underlying a sentence prior to the operation of any syntactic transformations. This structure is the result of application of the phrase structure rules and lexical insertion transformations which insert items from the lexicon into the skeletal structure.[5] Chomsky argued in *Aspects* that this underlying structure, which he there

[5] In the theory in *Aspects*, grammatical relations like *subject* and *object* are read off the syntactic structure. However, the relations themselves are semantic, so *subject* means *understood subject of*. A decade later, the theory of Relational Grammar (Perlmutter 1980) turned this view on its head, maintaining that grammatical relations are the primitives of the grammar. In Relational Grammar, grammatical relations are purely structural relations. This means that grammatical relations can be altered by transformations, and the major Relational Grammar syntactic processes had just this purpose.

named deep structure, is the locus of important generalizations and constitutes a coherent level of representation. Let us say a bit more about the latter concept before we move on.

Levels of representation were introduced into the theory in the following way in *LSLT*:

> We define, in general linguistic theory, a system of levels of representation. A level of representation consists of elementary units (primes), an operation of concatenation by which strings of primes can be constructed, and various relations defined on primes, strings of primes, and sets and sequences of these strings. Among the abstract objects constructed on the level **L** are **L**-markers that are associated with sentences. The **L**-marker of a sentence *S* is the representation of *S* on the level **L**. A grammar of a language, then, will characterize the set of **L**-markers for each level **L** and will determine the assignment of **L**-markers to sentences (Chomsky 1975: 6).

The child learning a language is assumed to bring knowledge of the levels to bear on the task of learning. That is, the child must learn properties of the language at each level, but knows the levels in advance, hence, knows what to look for. The levels are part of Universal Grammar. Of course, the *linguist* does not know in advance of research what the levels are. Determining them is a scientific question, one of biological psychology. Throughout the years, Chomsky and others have devoted considerable attention to determining just what the levels of representation are in the human language faculty. In *LSLT*, the levels were considered to be phonetics, phonemics, word, syntactic category, morphemics, morphophonemics, phrase structure and transformations. Throughout the years, the levels have changed in important and interesting ways.

Chomsky's major arguments for the new level, deep structure, in *Aspects* were that it resulted in a simpler overall theory, and at the same time it explained the absence of certain kinds of derivations that seemed not to occur (or at least seemed not to be needed in the description of sentences of human languages). Taking the second of these points first, Chomsky argued that while there is extensive ordering among singulary transformations (situations where a derivation produces an unacceptable sentence if two transformations are applied in reverse order), "there are no known cases of ordering among generalized transformations although such ordering is permitted by the theory of Transformation-markers" (Chomsky 1965: 133) (see also Fillmore 1963, Lees 1963). Further, while there are many cases of singulary transformations that must apply to a constituent sentence before it is embedded, or that must apply to a matrix sentence after another sentence is embedded in it, "there are no really convincing cases of singulary transformations that must apply to a matrix sentence before a sentence transform is embedded in it" (Chomsky 1965: 133).

As for the first argument, Chomsky claimed that the theory of transformational grammar is simplified by this change, since the notions "generalized transformation" and "Transformation-marker" are eliminated entirely. The P-markers in the revised theory contain all of the information of those in the *LSLT* version, but they also indicate explicitly how the clauses are embedded in one another, that is, information that had been provided by the embedding transformations and T-markers.

This change in the theory of phrase structure, which has the effect of eliminating generalized transformations, also has consequences for the theory of singulary transformations. As indicated above, in the *Aspects* theory, as in *LSLT*, there is extensive ordering among singulary transformations. In both frameworks, the set of singulary transformations was seen as a linear sequence: an ordered list. Given the *Aspects* modification, this list of rules applies cyclically, first operating on the most deeply embedded clause, then the next most deeply embedded, and so on, working up the tree until they apply on the highest clause, the entire generalized P-marker. Thus, singulary transformations apply to constituent sentences "before" they are embedded, and to matrix sentences "after" embedding has taken place. "The ordering possibilities that are permitted by the theory of Transformational-markers but apparently never put to use are now excluded in principle" (Chomsky 1965: 135).

3. The syntax/semantics interface in early generative grammar and beyond

An important question for any syntactic theory is how syntax relates to semantics: what the precise connection is between form and meaning. In *LSLT*, the T-marker contains all of the structural information relevant to semantic interpretation. Katz and Postal (1964) proposed a severe restriction on just how this structural information could be accessed. In particular, they postulated that the only contribution of transformations to semantic interpretation is that they interrelate P-markers. The slogan at the time was that "transformations do not change meaning". As Chomsky put it, (generalized) transformations combine semantic interpretation of already interpreted P-markers in a fixed way. In the revised theory, which Chomsky called the Standard Theory, the initial P-marker, now a deep structure, then contains just the information relevant to semantic interpretation. To summarize the model:

> … the syntactic component consists of a base that generates deep structures and a transformational part that maps them into surface structures. The deep structure of a sentence is submitted to the semantic component for semantic interpretation, and its surface structure enters the phonological component and undergoes phonetic interpretation. The final effect of a grammar, then, is to relate a semantic interpretation to a phonetic representation – that is, to state how a sentence is interpreted. (Chomsky 1965: 135-136)

To carry out this program, Chomsky (1965) adopted the proposal of Katz and Postal (1964) that many seemingly "meaning-changing" optional transformations of *LSLT* be replaced by obligatory transformations triggered by a marker in the deep structure. To take one example, earlier we noted that in *LSLT*, simple questions and the corresponding statements are derived from the same initial P-marker. In the revision, those initial P-markers would be very similar but not identical. The former would contain a marker of interrogation that would both signal the difference in meaning and trigger the inversion that results in the auxiliary verb appearing at the front of the sentence. Katz and Postal also noted that there are languages such as Japanese in which the Q-marker is spelled out as a separate morpheme.

At this point in the development of the theory, the model can be graphically represented as follows, with deep structure doing the semantic work formerly done by the T-marker:

(13) Deep Structure ⇒ Semantic Interpretation
⇓
Transformations
(operating cyclically)
⇓
Surface Structure ⇒ Phonetic Interpretation (via the "sound-related" levels of Morphophonemics, phonemics, and phonetics)

Some researchers soon challenged this framework. Generative Semantics built on the work by Katz and Postal (1964), and especially the claim that deep structure determines meaning (Lakoff 1971). For generative semantics, syntax is not the primary generative component. Rather, each meaning is represented by a different deepest representation (much more abstract that Chomsky's deep structure). On this view, transformations can, and often must, be far more complex and powerful than those in the *Aspects* model. There was intense debate about these issues in the late 1960s and into the 1970s before Generative Semantics largely disappeared from the scene, partly because the main practitioners came to develop different interests. However, central aspects of Generative Semantics have survived in different contemporary frameworks such as Cognitive Linguistics, Construction Grammar, and generative grammar including Chomskyan approaches. For example, Generative Semantics assumed that causative structures have a cause morpheme in the syntax, which is an approach that is found in recent work (see e.g., Harley 1995). Baker's (1988) Uniformity of Theta Assignment Hypothesis (UTAH), which states that identical thematic relationships are represented by identical structural relationships, is, in essence, another example of a proposal from Generative Semantics that has returned. Yet another, which we will discuss below, is the elimination of deep structure as a level of representation.

Let us now return to the chronological history. By the time *Aspects* was published, there were already questions about initial structure as the sole locus of semantic interpretation. To take just one example, Chomsky (1957) observed that in sentences with quantifiers (see chapter 22), the derived structure has truth conditional consequences. (14a) may be true while (14b) is false, for instance if one person in the room knows only French and German, and another only Spanish and Italian (see also chapter 3, ex. (13)).

(14) a. Everyone in the room knows at least two languages.
b. At least two languages are known by everyone in the room.

In the theory of Chomsky (1957), this is not problematic, since semantic interpretation is based on the T-marker. However, in the *Aspects* framework, there is a problem, as Chomsky acknowledges. He speculates that the interpretive difference between (14a) and (14b) might follow from discourse properties, rather than grammatical ones. The general problem, though, came to loom larger and larger, leading to a theory in which both deep structure and surface structure contribute to semantic interpretation. The core idea was introduced by Jackendoff (1969) and then elaborated by Chomsky (1970a) (see also e.g., Bach 1964, McCawley 1968), and it is clearly different from the view held by generative semantics. In this so-called Extended Standard Theory the contribution of deep structure concerns "grammatical relations" such as understood subject and object of (cf. fn. 5). The contribution of surface structure concerns virtually all other aspects of meaning, including scope, as in the examples mentioned just above, anaphora, focus and presupposition.

Alongside these questions about deep structure as the sole locus of semantic interpretation, there were also challenges to its very existence. Postal (1972) argued that the best theory is the simplest, which, by his reasoning, included a uniform set of rules from semantic structure all the way to surface form, with no significant level (i.e., deep structure) in between. And McCawley (1968) explicitly formulated an argument against deep structure on the model of Morris Halle's (1959) famous argument against a level of taxonomic phonemics. McCawley's argument is based on the interpretation of sentences with *respectively*, such as (15).

(15) Those men love Mary and Alice respectively.

McCawley argues that a *respectively*-transformation relates (16) to (15).

(16) that man (x) loves Mary and that man (y) loves Alice

For McCawley, this is a syntactic operation since it involves conjunction reduction. McCawley then notes that there is a corresponding semantic relation between (17) and (18).

(17) $\forall x{:}x \in M$ [x loves x's wife]
(18) These men love their respective wives

For generative semanticists, such as McCawley, since there is no syntactic level of deep structure, there is no a priori need to separate the two operations involved in (15)-(16) and (17)-(18). The deepest level of representation is a semantic representation. But in a theory with deep structure, the syntactic operation involved in (15)-(16) would necessarily be post-deep structure, while the operation implicated in (17)-(18) would necessarily be in a different module, one linking a syntactic representation with a semantic representation. Purportedly, then, a generalization is missed, as in Halle's classic argument.

Chomsky (1970b) considers this argument, but rejects it, claiming that it rests on an equivocation about exactly what the relevant rule(s) would be in the theories in question. Chomsky points out that it is possible to give a more abstract characterization of the transformations such that one is not syntactic and the other is not semantic. Therefore there is no argument against deep structure here. Chomsky does, however, accept McCawley's contention that it is necessary to provide justification for the postulation of deep structure. But he observes that the same is true of surface structure or phonetic representation, or, in fact, any theoretical construct. How can such justification be provided?

> There is only one way to provide some justification for a concept that is defined in terms of some general theory, namely, to show that the theory provides revealing explanations for an interesting range of phenomena and that the concept in question plays a role in these explanations. (Chomsky 1970b: 64)

As far as Chomsky was concerned, this burden had been met, especially by the *Aspects* analysis of the transformational ordering constraints discussed above.[6]

One small simplification in the Extended Standard Theory model was the result of a technical revision concerning how movement transformations operate (Wasow 1972, Chomsky 1973, Fiengo 1974, 1977). Trace theory proposed that when an item moves, it leaves behind a 'trace', a silent placeholder marking the position from which movement took place. The motivation for this was that in important respects, movement gaps behave like positions that are lexically filled, an argument first made in Wasow (1972) and Chomsky (1973). Under trace theory, the importance of D-structure for semantic interpretation is reduced, and ultimately eliminated. Once S-structure is enriched with traces, even grammatical relations can be determined at that derived level of representation. Using the term LF ('Logical Form') for the syntactic representation that relates most directly to the interpretation of meaning and PF ('Phonetic Form') for the one relating most directly to how sentences sound, we have the so-called T-model in (19) (also called the (inverted) Y-model), which was at the core of Government and Binding theory.

```
(19)         D-structure
                  |
           Transformations
                  |
             S-structure
               /     \
             PF       LF
```

The precise nature of the connection between the syntactic derivation and semantic and phonological interfaces has been a central research question throughout the history of generative grammar. In the earliest generative model, the interface is the T-marker, which includes all of the syntactic structures created in the course of the derivation. Subsequent models had the following interfaces with semantics: The Standard Theory had D-structure, the Extended Standard Theory had D-structure and S-structure, whereas Government and Binding and early Minimalism had LF. Chomsky's most recent model even dispenses with LF as a level in the technical sense (Chomsky 2004). The Minimalist approach to structure building, where Merge is the basic operation, is much more similar to that of the 1950s than to any of the intervening models, which is to say that interpretation in the Minimalist model also could be more like that in the early

---

[6] It is worth noting that in *Aspects*, cyclicity and deep structure were intertwined. Later on, they were distinguished, which means that one has to reconsider the previous evidence for deep structure.

model, distributed over many structures. In the late 1960s and early 1970s, there were already occasional arguments for such a model from phonological interpretation as well as semantic interpretation. For example, Bresnan (1971) argued that the phonological rule responsible for assigning English sentences their intonation contour (see chapter 23) applies cyclically, following each cycle of transformations, rather than applying at the end of the entire syntactic derivation. There were similar proposals for semantic phenomena involving scope and anaphora put forward by Jackendoff (1972). Chomsky (2000, 2001, 2004) argued for a general instantiation of this distributed approach to phonological and semantic interpretation, based on ideas of Epstein (1999) and Uriagereka (1999), who called the approach 'Multiple Spell-Out'. Simplifying somewhat, at the end of each cycle (or 'phase' as it has been called for the past 10 years) the syntactic structure created thus far is encapsulated and sent off to the interface components for phonological and semantic interpretation. Thus, although there are still what might be called PF and LF components, there are no syntactic levels of PF and LF. Epstein argued that such a move represents a conceptual simplification, and both Uriagereka and Chomsky provided some empirical justification. We can view this conceptual simplification similarly to the elimination of D-structure and S-structure. Chomsky (1993) argued that both D-structure and S-structure should be dispensed with. Both levels are theory-internal, highly abstract and they are not motivated by conceptual necessity, as the semantic and phonological interfaces to a much greater extent are. Another way to put this is to say that the motivation for D-structure and S-structure is empirical. Chomsky argued that, contrary to appearances, it is possible to cover the same or even more empirical ground without postulating either S-structure or D-structure.[7]

The role of syntactic derivation becomes even more central on this view because there are no levels of representation at all. The syntax interfaces directly with sound and meaning.

4. The development of phrase structure

In this section, we will provide a history of the development of phrase structure (see also Fukui 2001 and chapters 4, 6, 7 and 8). We will start with a brief recap of PS grammars and then move on to different versions of X-bar theory. Lastly we will discuss the approach to phrase structure within the Minimalist Program: Bare Phrase Structure. Our focus throughout will mainly be on the Chomskyan versions of phrase structure; but we will also mention where other theories developed and why they developed.

4.1. Phrase structure grammars

Chomsky (1955, 1957) developed a theory of phrase structure which made use of context-free PS grammars ([Σ, F] grammars). In addition, the theory was based on derivations and equivalence classes of such derivations.

Chomsky (1957: 27-29, 87) defines phrase structure set-theoretically as in (20).

(20) Given a particular [Σ, F] grammar and a particular terminal string (i.e., string of terminal symbols):

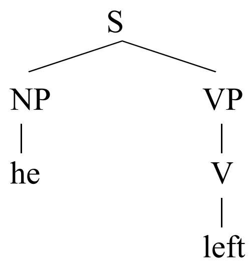
[7] See Section 2.7 below for discussion of an approach to cyclicity that, unlike that in *Aspects*, does not require recursion in the base (hence D-structure). In effect, this addresses one of the major *Aspects* arguments for D-structure.

a. Construct all of the equivalent PS derivations of the terminal string.
b. Collect all of the lines occurring in any of those equivalent derivations into a set. This set is the *phrase marker* (PM), a representation of the phrase structure of the terminal string.

The purpose of a PM is to tell us for each portion of the terminal string whether that portion comprises a constituent or not, and, when it comprises a constituent, what the "name" of that constituent is. Chomsky makes the following empirical claim: All and only what we need a PM to do is to tell us the "is a" relations between portions of the terminal strings and nonterminal symbols. Anything that tells us those and only those is a perfectly adequate PM; anything that does not is inadequate as a PM.

The PS rules can generate a graph-theoretic representation like the one in (21) (see Lasnik 2000: 29ff. for an illustration of how this works).

(21)



The tree tells us everything we have established concerning the "is a" relations. Note, however, that the tree encodes information that goes beyond the "is a" relations. The tree tells us that a VP is rewritten as V, and that the V is rewritten as *left*. It is an empirical question whether we need this additional information or not, say, for phonological, semantic, or further syntactic operations. If we do, then this particular set-theoretic model has to be rejected. If we do not, then the model is accepted, since we would like the minimal theory that does what has to be done. We will see later that the field has typically assumed that the set-theoretic model needs to be enriched in various ways.

Lasnik and Kupin (1977) showed that the algorithm for computing "is a" relations needs recourse only to the terminal string and the other members of the PM that consist of exactly one non-terminal symbol surrounded by any number of terminal symbols (what Lasnik and Kupin called monostrings). Hence Lasnik and Kupin proposed a construct called a *reduced phrase marker*, which includes only the terminal strings and the monostrings. See Lasnik (2000: section 1.2.6.1) for more discussion.

4.2. X-bar theory

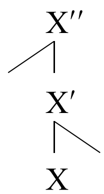
One problem in *LSLT* and *Syntactic Structures* is that the theory developed there allows PS rules like (23) alongside ones like (22) (Lyons 1968).

(22) NP → … N …
(23) VP → … N …

But there do not seem to be rules like (23). Why is this? The formalism allows both rules, and the evaluation metric (Chomsky 1965) judges them equally costly. Chomsky (1970a) was an attempt to come to grips with this problem. There it is proposed that there are no individual PS rules of the sort that did so much work in *Syntactic Structures* and even in *Aspects*. Rather, there is what is now known as the X-bar schema (see also chapter 11). X is a variable, ranging over category names such as V, N, and so on.

Here is the version of X-bar theory that Chomsky (1970a) presented (see also Emonds 1976 and Jackendoff 1977 for much relevant discussion).

(24) X′ → … X …
X″ → … X′ …

(25)


X′ and X″ are true complex symbols. Keep in mind that in *Syntactic Structures* NP looked like it had something to do with N, but in that system it really did not. NP was just one symbol that was written for mnemonic purposes with two letters. In X-bar theory, a category label is a letter plus a number of bars (originally written as overbars – e.g., $\bar{\text{X}}$ – but later written as primes – e.g., X′ – for typographical convenience). It can be thought of as an ordered pair. X is <X, 0>, X′ is <X, 1>, and X″ is <X, 2>. X-bar theory immediately explains why there are no rules like (23). This is because phrases have heads, i.e., they are endocentric, which is to say that phrases are projections of heads.

Chomsky also introduced the relational notions complement and specifier. A complement is a sister to a head. He argued that the notion complement does not play any role in transformations (Chomsky 1970a: 210), that is, complements cannot be the target qua complements of any transformational operations. At this point, there were general rules like (29) that subsumed rules like the ones in (26)-(28).

(26) NP → N Comp
(27) VP → V Comp
(28) AP → A Comp
(29) Comp → NP, S, NP S, NP Prep-P, Prep-P Prep-P, etc.

The rules in (29) should instead be replaced with the rule in (30).

(30) X′ → … X …

The dots in (30) indicate that there are no restrictions on what can be a complement and where the complement is placed vis-à-vis the head.

Chomsky then proposes that in order to “introduce further terminological uniformity, let us refer to the phrase associated with N′, A′, V′ in the base structure as the ‘specifier’ of these elements” (Chomsky 1970a: 210).

(31) X″ → [Spec, X′] X′

On this view, a specifier encompasses a heterogeneous set as it contains a variety of pre-head elements like auxiliaries in SpecV′, determiners in SpecN′, adverbials in SpecV′ and degree modifiers in SpecA′. As Jackendoff (1977: 14) points out, it is not clear whether Chomsky considers the specifier to be a constituent or an abbreviation for a sequence of constituents, like Comp. The diagrams in Chomsky (1970) show specifiers as constituents. Jackendoff (1977) argues against specifiers being constituents whereas Hornstein (1977) defends the claim that they are. However, beyond being a constituent and bearing a geometrical relation to a head, it is not clear what the defining characteristics of a specifier are (see also George 1980: 17).

Later a biconditional version of X-bar theory was developed, namely that phrases have heads, and heads project. Whenever a structure has an XP, it has an X (this is what Chomsky 1970a proposed), and whenever a structure has an X, it has an XP.

In Chomsky (1970a), the initial rule of the base grammar is as in (32).

(32) $S \rightarrow N'' \; V''$

This means that X-bar theory is not fully general: S and S′ (the latter the larger clause including a sentence introducing complementizer like *that*) do not fit into the theory in any neat way.[8] These labels are not projections of heads, unlike the other labels in the system. However, it is worth bearing in mind that Bresnan (1970) suggests that complementizers are essentially specifiers of sentences through the rule in (33).

(33) $S' \rightarrow \text{Comp} \; S$

This is in line with the general approach to specifiers during the 1970s, as complementizers here are analyzed on a par with auxiliaries, which were also specifiers.

It may be worth pausing to reflect on what pushed Chomsky to create X-bar theory.

> The development of X′ theory in the late 1960s was an early stage in the effort to resolve the tension between explanatory and descriptive adequacy. A first step was to separate the lexicon from the computations, thus eliminating a serious redundancy between lexical properties and phrase structure rules and allowing the latter to be reduced to the simplest (context-)free form. X′ theory sought to eliminate such rules altogether, leaving only the general X′ theoretic format of UG. The problem addressed in subsequent work was to determine that format, but it was assumed that phrase structure rules themselves should be eliminable. (Chomsky 1995a: 61)

The attempt was to do away with redundancies in favor of larger generalizations. Another way to say this is that when we impose strict constraints, the PS rules themselves vanish. It is possible to view the change from phrase structure rules to X-bar theory in the same way as Chomsky's (1973) generalization of some of Ross's (1967) locality "island"

[8] Though see Jackendoff (1977) for a way to in part solve this problem by identifying S with V′′ in his system. See also Hornstein (1977) who argues that S should be excluded from the X-bar convention.

constraints on movement (see chapter 18). In both cases, instead of more or less idiosyncratic properties, we get general properties that hold across categories. Baltin (1982: 2) puts the general development this way:

> The history of transformational generative grammar can be divided into two periods, which can be called expansion and retrenchment. During the early "expansion" period, a primary concern was the description of grammatical phenomena. […] The theory was correspondingly loose, and consequently failed to provide an adequate solution to the projection problem.[9] […] During the retrenchment period […] the focus of attention shifted from the construction of relatively complex […] statements to the construction of a general theory of grammar, restricted as to the devices it employed, which could be ascribed to universal grammar.

Chomsky (1970a) only discusses NPs, VPs and APs, not PPs. One goal of Jackendoff (1977) is to bring PPs under the X-bar theoretic fold. So at the end of the 1970s, a quite general picture of phrase structure had started to emerge.

Before we move on to the early Principles and Parameters view of phrase structure, it is worth considering a general problem that both Chomsky (1970) and Jackendoff (1977) face. The problem has been brought up most clearly by Stuurman (1985). Stuurman's goal is to defend what he calls "the single-projection-type hypothesis". Multiple projection types (X, X', X'', X^n), as assumed in Chomsky's and Jackendoff's works, are banned. Stuurman's thesis is that only one distinction is made internal to projections: the distinction between X^0 and X^1, or put differently, between a head and everything else. Stuurman argues that this provides a more restrictive phrase structure theory and a theory that is more easily learnable. Here is an example that he uses to make his claim.

In English, only the first hierarchical level projected from X^0 can dominate an NP.

(34) a. he [[met his wife] in Italy]
b. *he [[met in Italy] his wife]

Stuurman (1985: 8) points out that if we assume multiple projection-types, the facts in (34) can easily be captured directly at the level of PS as follows:

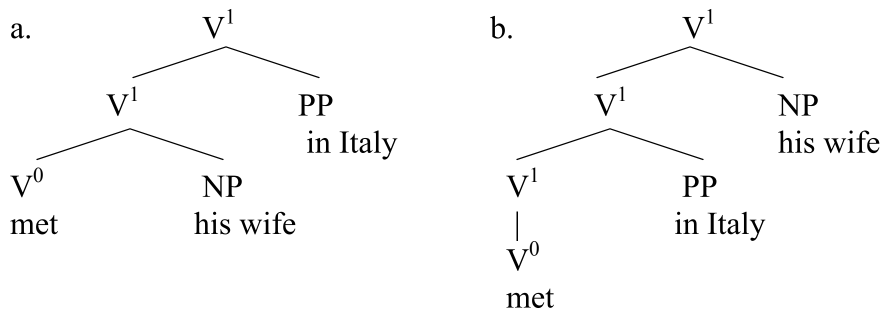
(35) a. $V^i \rightarrow \ldots V^j \ldots$, where $\ldots \neq$ NP,$\underline{i} > \underline{j} \geq 1$
b. $V^1 \rightarrow \ldots V^0 \ldots$, where $\ldots =$ NP,…

These restrictions are descriptively adequate, but as Stuurman stresses, they do not explain how a child can learn the distribution of NPs. Put differently, UG does not provide a rationale for why the constraints are the way they are: Why should UG not allow NP under V^i and exclude NP under V^1? Unless the rules in (35) are universal, children need access to negative data (i.e., that (34b) is bad), which they by assumption do not have access to.[10]

[9] That is, the problem of 'projecting' the correct grammar from limited input data.
[10] See also Stowell (1981: 71-75) for criticism based on arguments from acquisition.

Stuurman presents a different analysis where there is only one projection type. His theory, which we will not flesh out here, allows for both the structure in (36a) and (36b).

(36) a. b.



Here one needs an independent principle that filters out the structure in (36b). This structure has an NP that is not dominated by the first X^1 up from X^0. Stuurman argues that this filtering condition can be associated with an adjacency condition on Case theory, following Stowell (1981) (see chapter 16 for more discussion). That is, being a Case assigner is a lexical property, thus a property of X^0, not of X^1. (36b) is therefore ruled out independently of PS rules, as in Stowell's work.[11] Stuurman presents additional arguments for the single projection hypothesis. The point is that the view emerging in the late 1970s had important flaws, as it was too flexible and not principled enough. In the early 1980s, these flaws were addressed.

As research developed during the 1970s and 1980s, more and more of the elements that Chomsky and Jackendoff had analyzed as specifiers came to be analyzed as heads of particular functional projections (see also Abney 1987). As Chametzky (2000) points out, a notion of specifier emerged with the following characteristics: (i) typically an NP, (ii) it bears a certain relationship with the head. Stowell (1981: 70) summarizes the general characteristics of X-bar theory as follows:

(37) a. Every phrase is endocentric.
b. Specifiers appear at the XP-level; subcategorized complements appear within X′.
c. The head always appears adjacent to one boundary of X′.
d. The head term is one bar-level lower than the immediately dominating phrasal node.
e. Only maximal projections may appear as non-head terms within a phrase.

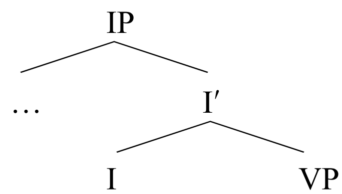
These were further developed during the Government and Binding era in the 1980s. Here we will focus on Chomsky (1986), since that work presents X-bar theory as it is best known.

Chomsky (1986, henceforth *Barriers*) provides a generalization of X-bar structure, though attempts had already been made in Chomsky (1981), Stowell (1981) and den Besten (1983), to mention the most important works. As we have seen, prior to

[11] In fact, Stowell (1981) argued for the general elimination of phrase structure rules, thus providing empirical motivation for the formalization of Lasnik and Kupin (1977).

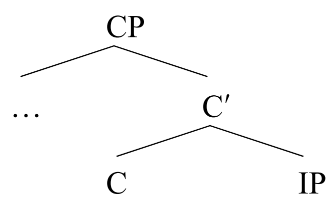
Barriers, the maximal projections were VP, NP, AP and PP. In addition, there was S, which gets rewritten as NP Infl VP, and S′, which gets rewritten as Comp S. Comp includes at least C and *wh*-expressions. The problem is that S does not conform to X-bar theory. It is not endocentric since it has no head, which means that there is no projection line from a head to a maximal projection. S′ is also not uniformly endocentric since when Comp is filled by phrasal material, it is not the head of S′. Because of these problems, Stowell (1981: chapter 6) suggests that the head of S is Infl, as illustrated in (38). This is very similar to Williams (1981: 251), who suggests that S is headed by Tense.

(38)



Once IP replaces S, a natural step is to reconsider S′. Stowell (1981: chapter 6) proposes that C is the head of S′. The optional specifier then becomes the target of *wh*-movement. We then have the structure in (39) (see also Chomsky 1986, and chapter 11, sect. 5).

(39)



With this in place, it is possible to formulate restrictions on movement based on what can appear in a head position and what can appear in a specifier position, cf. Travis (1984) and Rizzi (1990).

The reanalysis of S and S′ paves the way for a generalization of X-bar theory. Chomsky (1986: 3) proposes that X-bar theory has the general structure in (40), where X* stands for zero or more occurrences of some maximal projection and $X = X^0$.[12]

(40) a. $X' = X\ X''{*}$
b. $X'' = X''{*}\ X'$

Koizumi (1995: 137) argues that the traditional X-bar schema can be seen as expressing three claims, as given in (41).

(41) a. Asymmetry: A node is projected from only one of its daughters.
b. Binarity: A node may have at most two daughters.
c. Maximality: A head may project (at most) two non-minimal projections.

[12] This is what Chomsky said, but it is obviously not exactly what he meant. (40a) should read $X' = X\ Y''{*}$ because otherwise a verb, for example, can only take a VP complement, and similarly for (40b) and specifiers.

It should be mentioned that (40) does not force binarity since a node may have more than two daughters. One can either restrict X-bar theory so that it does observe binarity by hard-wiring it into the X-bar theory, or e.g., follow the proposal of Kayne (1984, 1994) that independent grammatical constraints require all branches in a tree to be binary (see below).

Chomsky (1986: 4) points out that specifiers are optional whereas the choice of complements is determined by the Projection Principle. The latter is a principle that says that representations at each syntactic level are projected from the lexicon.

Following up on the theory in *Barriers*, many researchers developed somewhat different versions of X-bar theory. Fukui and Speas (1986) claim that there are significant differences between lexical and functional projections, e.g., VP and IP. They argue that lexical categories may iterate specifiers as long as all these positions are fully licensed and can be interpreted at LF. Functional categories, on the other hand, only have one unique specifier position.[13]

Hoekstra (1991; see also Hoekstra 1994) argues that specifiers are stipulated in X-bar theory. Rather, Hoekstra argues, specifiers should be defined through agreement: A specifier always agrees with its head (see also chapter 17). Hoekstra also eliminates the phrase-structural distinction between adjuncts and specifiers and argues that an adjunct can be defined as an element that does not agree with the head of the projection it is adjoined to. Recently, several researchers have argued that specifiers are problematic and should not be part of phrase structure (Hoekstra 1991, Kayne 1994, Cormack 1999, Starke 2004, Jayaseelan 2008).

Kayne (1994) puts forward a novel theory of phrase structure. He suggests there is one universal order and that this order is as in (42).

(42) specifier > head > complement

Throughout the history of generative grammar, it had generally been an assumption that languages vary in their base structure. PS rules encode this directly as in (43) for an English VP and (44) for a Japanese VP.

(43) VP $\rightarrow$ V NP
(44) VP $\rightarrow$ NP V

In the Government and Binding era, a common analysis of this variation was given in terms of the head parameter. Contrary to these analyses, Kayne claims that linear and hierarchical order are much more tightly connected. He argues that the property of antisymmetry that the linear precedence ordering has is inherited by the hierarchical structure.[14] The Linear Correspondence Axiom is the basic property of phrase structure, and familiar X-bar theoretic properties follow from it.

[13] See also Stuurman (1985: 182) for a similar claim, though Stuurman claims that this also holds for lexical categories.

[14] We speculate that Kayne intended asymmetry rather than antisymmetry. An antisymmetric relation R is one where if $(a, b) \in R$ and $(b, a) \in R$, then $a = b$. Asymmetry is a stronger property: $(a, b) \in R \rightarrow (b, a) \notin R$. Since items evidently do not precede themselves, the weakening loophole of antisymmetry is not needed.

(45) *Linear Correspondence Axiom*

d(*A*) is a linear ordering of *T*. (Kayne 1994: 6)

The nonterminal-to-terminal dominance relation is represented by *d*. This relation d is a many-to-many mapping from nonterminals to terminals. For a given nonterminal X, d(X) is the set of terminals that X dominates. *A* is a set of ordered pairs $<X_j, Y_j>$ such that for each j, X_j asymmetrically c-commands Y_j. *A* contains all pairs of nonterminals such that the first asymmetrically c-commands the second, thus it is a maximal set. *T* stands for the set of terminals.

At this point, we will turn to a brief description of Bare Phrase Structure, which partly incorporates Kayne's ideas, since this is the current approach to phrase structure in Chomskyan generative grammar.

4.3. Bare Phrase Structure and Cartography

Kayne's theory forces the elimination of the distinction between X′ and XP since his linearization algorithm does not make this distinction. Chomsky (1995a, b) went further and argued that X-bar levels should be eliminated altogether. This is the theory of Bare Phrase Structure (BPS; see also section 2.4.2 of chapter 4).

The gist of BPS is summarized in the following quote: "Minimal and maximal projections must be determined from the structure in which they appear without any specific marking; as proposed by Muysken (1982) they are relational properties of categories, not inherent to them" (Chomsky 1995a: 61): "What I will propose is that bar level is not a primitive of the grammar at all, rather 'maximal projection' and 'minimal projection' are defined terms, and intermediate projections are simply the elsewhere case" (Muysken 1982).[15] Chomsky (1995b: 242) tied this to the Inclusiveness Condition, which bans any marking of maximal and minimal projections.[16]

(46) *Inclusiveness Condition*

Any structure formed by the computation is constituted of elements already present in the lexical items. No new objects are added in the course of computation apart from rearrangements of lexical properties. (Chomsky 1995b: 228)

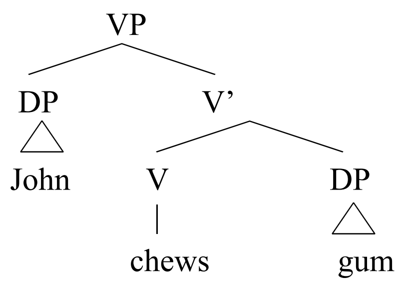
Another way to look at BPS is to say that phrase structure consists solely of lexical items. No extrinsic marking is necessary. This means that instead of a phrase like (47), phrases look like (48). Here we are setting aside how verbs get their inflection and where the arguments really belong in the structure – the important point at hand is the difference between the two structures.

---

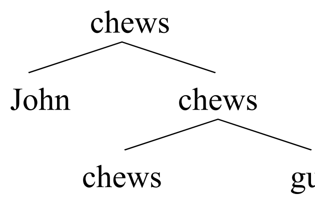
[15] This way of looking at phrase structure is closely related to Speas (1990: 35).

[16] This condition is an obvious extension of an idea in Katz and Postal (1964: 44-45), further developed in Chomsky (1965: 132) when he suggests that transformations cannot introduce meaning-bearing elements.

(47)



(48)



These lexical items are accessed at the LF interface. No units apart from the lexical items can be part of the computation. Thus bar-levels have no existence within BPS. For a critical discussion of some problems with BPS, see Starke (2004) and Jayaseelan (2008).

Shortly after BPS had been developed in Chomsky (1995a, b), Rizzi (1997) initiated what has become known as the cartographic approach. This approach assumes an expansion of functional structure, an expansion that is claimed to be necessary on empirical grounds. See chapter 12 for discussion.

This concludes our rather brief overview of the history of phrase structure. A common thread has been the reduction and generalization that started with Chomsky (1955). X-bar theory was a generalization of PS grammars but at the same time a reduction in that the core primitives of the theory were fewer. Chomsky (1986) also made significant generalizations of the X-bar theory in Chomsky (1970a). Lastly, BPS has provided the last reduction that we have seen so far, where even the existence of bar levels is denied.

5. Rules and filters versus principles

Most of the early work in generative syntax was done on English. A few important exceptions were Kuroda (1965), Matthews (1965), Ross (1967), Perlmutter (1968) and Kayne (1969). However, especially with the publication of Kayne (1975), it became more and more common to investigate different languages.[17] Kayne gave a range of different language-particular rules for French and in many cases compared them to the syntax of English. Slightly later, Jaeggli (1980) and Rizzi (1982) conducted in-depth studies of other Romance languages. Crucially, though, this enterprise centered on formulating language-specific and construction-specific rules, and what may be universal across languages was not given as much attention.

Chomsky and Lasnik (1977) pointed out that early work in pursuit of descriptive adequacy led to an extremely rich theory of transformational grammar. For a formalization that encompasses much descriptive practice, see Peters and Ritchie (1973). Even this extremely rich theory does not encompass such devices as structure-building

[17] This also happened as textbooks on the *Syntactic Structures* and *Aspects* frameworks were written.

rules, global rules, transderivational constraints, and others that had often been proposed. Let us take a quick look at global rules and transderivational constraints.

A global rule is a rule that state conditions on “configurations of corresponding nodes in non-adjacent trees in a derivation” (Lakoff 1970: 628). Thus, global rules go far beyond the usual Markovian property of transformational derivations. An example of a global rule is provided by Ross (1969). Ross observed that the island constraints on movement he proposed in Ross (1967) only hold if the island-forming node is present in surface structure. The constraints do not hold, however, if a transformation (“Sluicing” in this case; see chapter 19) subsequently deletes that node. An example illustrating this is given in (49)-(50).

(49) *Irv and someone were dancing, but I don’t know who Irv and were dancing.
(50) Irv and someone were dancing, but I don’t know who.

The conclusion drawn from this is that island constraints cannot just mention the point in the derivation at which the movement rule applies, nor just the surface structure. The constraints must mention both.

As for transderivational constraints, these are constraints that depend on properties of derivations other than the one currently being constructed. Hankamer (1973) argues for transderivational constraints based on a detailed analysis of Gapping (see chapter 19). Among others, he considers the data in (51)-(54) (Hankamer 1973: 26-27).

(51) Max wanted Ted to persuade Alex to get lost, and Walt, Ira.
(52) … and Walt *[wanted] Ira [to persuade Alex to get lost]
(53) … and Walt *[wanted Ted to persuade] Ira [to get lost]
(54) … and [Max wanted] Walt [to persuade] Ira [to get lost]

In order to block Gapping in (52)-(53), Hankamer argues that a constraint is needed that makes reference to other structures that might have been created, even from different deep structures. In particular, the reason (51) cannot be derived from (52) or (53) is that it *can* be derived from (54). Space considerations prevent us from elaborating further, though we should acknowledge that Hankamer suggests that the constraint at issue here is universal, thus raising no learnability concerns.

Returning to our main discussion, any enrichment of linguistic theory that extends the class of possible grammars requires strong empirical motivation. This, Chomsky and Lasnik (1977) argued, is generally missing in the case of devices that exceed the framework of Chomsky (1955), Peters and Ritchie (1973), and comparable work; cf. Dougherty (1973), Chomsky (1972), Brame (1976).

Note that the work of Chomsky and many others has consistently tried to reduce the descriptive power of the transformational component. The framework in *Aspects* is more restricted than the one in *LSLT*, and Chomsky (1973) is much more restricted than *Aspects*. In the 1980s, many researchers argued that we should make transformations as general as *Move Alpha*, or even *Affect Alpha*, as in Lasnik and Saito (1984, 1992).

Chomsky and Lasnik (1977) contributed to these developments by proposing a framework that attempted to restrict the options that are available in this narrower, but still overly permissive framework, so that it is possible to approach one of the basic goals of linguistic theory: to provide, in the sense of *Aspects*, explanations rather than descriptions and thus to account for the attainment of grammatical competence. They assumed that Universal Grammar is not an "undifferentiated" system, but rather a system that incorporates something analogous to a theory of markedness. Specifically, there is a theory of core grammar with highly restricted options, limited expressive power, and a few parameters. Systems that fall within core grammar constitute the unmarked case; one can think of them as optimal in terms of the evaluation metric. An actual language is determined by fixing the parameters of core grammar and then adding rules or rule conditions, using much richer resources, perhaps resources as rich as those contemplated in the earlier theories of transformational grammar noted above.

Filters were supposed to bear the burden of accounting for constraints, which, in the earlier and far richer theory, were expressed in statements of ordering and obligatoriness, as well as contextual dependencies that cannot be formulated in the narrower framework of core grammar. The hypothesis in Chomsky and Lasnik (1977) was that the consequences of ordering, obligatoriness, and contextual dependency could be captured in terms of surface filters. Furthermore, they argued that the above-mentioned properties could be expressed in a natural way as surface filters that are universal, or else the unmarked case.

We see that the idea of a distinction between parameters and principles is already present in Chomsky and Lasnik (1977). However, in this framework, there are only a few parameters that affect the core grammar. Besides these parameters, there are a number of language-specific rules. An example is the filter in (55) that blocks *for-to* constructions in Standard English.

(55) *[for-to]
(56) *We want for to win.

As Chomsky and Lasnik (1977: 442) point out, this filter is a "dialect" filter, meaning that it is not a principle of Universal Grammar. They discuss a range of filters, and some of them are like (55) in being outside of core grammar, whereas others, like the Stranded Affix filter of Lasnik (1981), are argued to be part of Universal Grammar.

With Chomsky (1981) the conception of rules and filters changed somewhat. The part related to rules stayed intact, since there is no distinction between rules and principles. Both are assumed to be universal and part of Universal Grammar. But instead of filters that can both be language- and construction-specific, Chomsky suggested that we should conceive of variation in terms of parameters (hence the name *Principles and Parameters Theory*; see chapter 4). The following quote brings out the main difference.

> If these parameters are embedded in a theory of UG that is sufficiently rich in structure, then the languages that are determined by fixing their values one way or another will appear to be quite diverse (Chomsky 1981: 4).

The parameters are assumed to be part of UG and together they should both yield the variation we observe and an answer to Plato's problem: How do we know so much given the limited evidence available to us? In the realm of language, the question is how the child can arrive so rapidly at its target grammar given the input it gets. An important part of the theory was that parameters were supposed to represent clusters of properties: "[I]deally we hope to find that complexes of properties […] are reducible to a single parameter, fixed in one or another way" (Chomsky 1981: 6). Rizzi (1982) gave a nice example of this when he argued that there are correlations between thematic null subjects, null expletives, free inversion and *that*-trace effects (**Who do you think that __ won the race*).

This model was therefore a sharp break from earlier approaches, under which universal grammar specified an infinite array of possible grammars, and explanatory adequacy required a presumably unfeasible search procedure to find the highest-valued one, given primary linguistic data. The Principles and Parameters approach eliminated all this. There is no enumeration of the array of possible grammars. There are only finitely many targets for acquisition, and no search procedure apart from valuing parameters. This cut through an impasse: Descriptive adequacy requires rich and varied grammars, hence unfeasible search; explanatory adequacy requires feasible search.

See chapters 4, 24 and 25 for further discussion of parameters.

6. Derivations

The general issue of derivational versus representational approaches to syntax has received considerable attention throughout the history of generative grammar. A derivational approach argues that there are constraints on the processes by which well-formed expressions are generated whereas a representational approach argues that there is a system of well-formedness constraints that apply to structured expressions (see Frank 2002 for more discussion of this general issue). Internally to the major derivational approach, transformational grammar, a related issue arises: Are well-formedness conditions imposed specifically at the particular levels of representations made available in the theory, or are they imposed "internal" to the derivation leading to those levels?[18] Like the first question concerning whether derivations exist, it is a subtle one, perhaps even subtler than the first, but since Chomsky (1973), there has been increasing investigation of it, and important arguments and evidence have been brought to bear (see Freidin 1978 and Koster 1978 for illuminating early discussion).

However, generative theories disagree on whether derivations actually exist or not. Typically this disagreement emerges when the question of whether there are transformations is considered, since this is the main case where one can impose derivational constraints. Any phrase structure representation has to be generated somehow, and one can arguably claim that the generation of such a tree is derivational. This is not where the disagreement lies; rather, it concerns whether one can impose constraints on derivations or not. Chomskyan generative grammar, especially since the very important work of Ross (1967), has always assumed that this is possible and that it is a virtue of the theory. However, let us consider some non-transformational theories (see also Frank 2002 for useful discussion, and Harman 1963 for a very early formulation of a non-transformational generative theory). Most of these developed in the wake of Chomsky's (1973, 1977) theorizing based on the important discoveries in Ross (1967).

---

[18] There are of course hybrid theories as well. Chomsky (1981), for example, proposes well-formedness conditions on deep structure, on surface structure, and on the application of transformations between grammatical levels.

Lexical-Functional Grammar (LFG) (Kaplan and Bresnan 1982, Bresnan 2001) eliminates transformations and increases the role of structural composition. This is a theory where the lexical expressions are of crucial importance. LFG argues that lexical representations have a richer hierarchical structure than in the Chomskyan theory. The theory also assumes parallel levels of representation: constituent structure, functional structure and argument structure all constitute independent levels of representation. Since the theory does not have transformations, dependencies are established by interaction between the different levels and by lexical entries that have been transformed by lexical rules. For example, an analysis of the passive assumes that there are two lexical entries of the verb in the lexicon and that there are linkages that determine the appropriate thematic dependencies. See chapter 6 for more discussion of LFG.

Generalized Phrase Structure Grammar (GPSG) (Gazdar, Klein, Pullum and Sag 1985) eliminates transformations in a different way. In this theory, a derivation consists of context-free phrase structure rules. Metarules that modify the phrase structure rules are used to established dependencies in a way reminiscent of Harman 1963. This is to say that *wh*-movement, for example, is captured through additional phrase structure rules. Chapter 7 discusses GPSG in detail.

As Frank (2002: 8) points out, all these non-transformational theories share with transformational theories the property that there are no privileged intermediate levels of syntactic structure. This has been the case since Chomsky (1965), but it was not true of Chomsky (1955, 1957), where kernel structures constituted such intermediate structures. Put differently, something needs to prevent nonlocal dependencies from being created. However, a non-transformational theory that returns to a theory that is closer to that of Chomsky (1955) is Tree Adjoining Grammar (Joshi, Levy and Takahashi 1975, Joshi 1985). We briefly described this theory in section 2.1 above; see also chapter 8.

In theories of the Chomskyan sort, based on transformational movement operations, a question arises: What determines whether movement occurs? In the Move α framework, all such processes were completely free (see e.g., Lasnik and Saito 1992 for a detailed version of this theory). There were no triggers; rather, there were representational constraints that had to be satisfied for a structure to be convergent. Even though representationalist approaches have been developed in recent years (see in particular Brody 1995, 2002, 2003), Chomsky and most researchers within Chomskyan generative grammar have defended a derivationalist approach where movement is triggered.[19] Chomsky (1995) argues on conceptual and, to some extent, empirical grounds that movement is always morphologically driven. That is, there is some formal feature that needs to be checked, and movement provides the configuration in which the checking can take place. Chomsky also provides reasons that, all else being equal, covert movement (movement in the LF component) is preferred to overt movement, a preference that Chomsky calls "Procrastinate". When movement is overt, rather than covert, then, it must have been forced to operate early by some special requirement. The major phenomenon that Chomsky considers in these terms is verb raising, following the influential work of Pollock (1989). He also hints at a contrast in object shift, overt in some languages and covert in others. Chomsky (1993, 1995a, 1995b) codes the driving

[19] Almost from the earliest days of generative grammar, there were qualms about optional transformations: "An obvious decision is to consider minimization of the optional part of the grammar to be the major factor in reducing complexity" (Chomsky 1958/1962: 154).

force for overt movement into strong features, and presents three successive distinct theories of precisely how strong features drive overt movement. These three theories, which we will summarize immediately, are of interest to our question, since the first two of them are explicitly representational in the relevant sense, while the third is derivational.

(57) a. A strong feature that is not checked in overt syntax causes a derivation to crash at PF. (Chomsky 1993)
b. A strong feature that is not checked (and eliminated) in overt syntax causes a derivation to crash at LF. (Chomsky 1995a)
c. A strong feature must be eliminated (almost) immediately upon its introduction into the phrase marker. (Chomsky 1995b)

All three of these proposals are designed to force overt movement in the relevant instances (e.g., verb raising in French where a strong V feature of Infl will cause a violation in one of the three ways listed in (57) if overt movement does not take place) and all are framed within a Minimalist conception of grammar. The work of building structure is done by generalized transformations, as it was before recursion in the base was introduced in Chomsky (1965). This return to an earlier approach replaces a partly representational view with a strongly derivational one.

Chomsky (1993) argues that the treatment in (57a) follows from the fact that parametric differences in movement, like other parametric differences, must be based on morphological properties reflected at PF. (57a) makes this explicit. Chomsky suggests two possible implementations of the approach:

> … "strong" features are visible at PF and "weak" features invisible at PF. These features are not legitimate objects at PF; they are not proper components of phonetic matrixes. Therefore, if a strong feature remains after Spell-Out, the derivation crashes… Alternatively, weak features are deleted in the PF component so that PF rules can apply to the phonological matrix that remains; strong features are not deleted so that PF rules do not apply, causing the derivation to crash at PF (Chomsky 1993: 198).

There is presumably only one other possible type of representational approach, given minimalist assumptions: one that involves LF, rather than PF. Chomsky (1995a) proposes such an analysis, (57b), based on an empirical shortcoming of (57a). What is at issue is the unacceptability of sentences like (58).

(58) *John read what?

Assuming that the strong feature forcing overt *wh*-movement in English resides in interrogative C,[20] the potential concern is that C, since it has no phonetic features, might be introduced in the LF-component, where, checked or not, it could not possibly cause a

[20] Notice that in English, the relevant strong feature could not reside in the *wh*-phrase, since in multiple interrogation, all but one of the *wh*s remain in situ, hence unchecked in overt syntax:
(i) Who gave what to who?

PF crash, since it has no phonetic features, and therefore as far as PF knows, the item does not exist at all. Yet (58) is bad as a non-echo question, so such a derivation must be blocked. This problem arises in the general context of fitting lexical insertion into the grammar. In most circumstances, there is no need for a specific prohibition against accessing the lexicon in the PF or LF component. (58) represents a rare problem for the assumption that lexical insertion is free to apply anywhere. Chomsky (1995a: 60-61) suggests that he root C head has a feature that requires overt *wh*-movement. Unless this feature is checked prior to Spell-Out, the derivation will crash at LF. Chomsky proposes to implement this basic idea in the following way: "Slightly adjusting the account in Chomsky (1993), we now say that a checked strong feature will be stripped away by Spell-Out, but is otherwise ineliminable" (Chomsky 1995a: 61).

Chomsky (1995b) rejects the representational approach in (57a), and the conceptual argument he gives evidently applies equally to the alternative representational approach in (57b). He discounts such an account as an evasion, and proposes what he claims is a more straightforward statement of the phenomenon:

> … formulation of strength in terms of PF convergence is a restatement of the basic property, not a true explanation. In fact, there seems to be no way to improve upon the bare statement of the properties of strength. Suppose, then, that we put an end to evasion and simply define a strong feature as one that a derivation "cannot tolerate": a derivation D → Σ is canceled if Σ contains a strong feature. (Chomsky 1995b: 233)

In summary, strong features trigger a rule that eliminates them. This approach is strongly derivational. There are problems with this account (see Lasnik 2001 for detailed discussion), but the goal here has merely been to outline the ways one can think of the trigger question in either derivational or representational terms.

Since Chomsky (1995b), the assumption is that movement is triggered by feature checking. But while feature checking was originally thought to be possible only in specific derived configurations (the Spec-head relation and head-adjunction configurations, in particular), in more recent work it is contingent merely on the establishment of an Agree relationship between a c-commanding Probe and a Goal. The introduction of the Agree mechanism divorces the movement trigger from agreement, contrary to the framework in Chomsky (1993) where elements moved to specifiers to undergo agreement with a head (see chapter 17 for discussion). However, even if features have to be checked, it is not clear that the approach is fully derivational. The typical assumption is that a derivation crashes unless all features are checked prior to the interfaces, which in effect is a representational condition based on features. However, the operations defined on features are derivational as they unfold as the structure is being built and they are limited by grammatical principles (e.g., intervention effects or the Phase Impenetrability Condition; see Chomsky 2001 and chapters 4 and 18 for discussion). Therefore it seems valid to say that there are both derivational and representational aspects and that both play important roles in grammar in this model.

7. The advent of economy principles in Principles and Parameters Theory

As we have seen, a major minimalist concern involves the driving force for syntactic movement. From its inception in the early 1990s, Minimalism has insisted on the last-resort nature of movement: in line with the leading idea of economy, movement must happen for a reason and, in particular, a formal reason. The Case Filter, which was a central component of the Government and Binding theory was thought to provide one such driving force. Chapter 17 illustrates this at length, so we will not discuss it here. Instead we will offer two other examples: Relativized Minimality and the Extension Condition.

An important instance of economy is what Luigi Rizzi (1990) called Relativized Minimality (see chapter 18 for more discussion). Chomsky and Lasnik (1993) reinterpreted Rizzi's groundbreaking work in terms of least effort. Let us illustrate that here by way of a phenomenon called Superiority, which has often been analyzed as a Relativized Minimality effect. Consider the following examples:

(59) Guess who bought what?
(60) *Guess what who bought?

In this situation, there might seem to be an option. One could either front *who* or *what*. As (59) and (60) show, only the former is licit. In such a situation, you always have to pick the closest to the position where the element ends up after moving, as first observed in something like these terms by Chomsky 1973. Put differently, one should minimize the distance traveled by the moving element, an instance of 'economy' of derivation.

Another potential example of an economy condition relates to the Extension Condition. This condition requires that a transformational operation extends the tree upwards. In Chomsky (1965) the requirement that derivations work their way up the tree monotonically was introduced, alongside D-structure. Generally this is known as the requirement of cyclicity. Chomsky used this to explain the absence of certain kinds of derivations, but also as an argument against generalized transformations and for D-structure. But it was cyclicity, rather than D-structure, that was crucial in the account. As we have discussed above, Minimalism rejects D-structure and reinstates generalized transformations, but it still preserves cyclicity, thus ruling out the anticyclic derivations that were the original concern. The Minimalist Extension Condition demands that both the movement of material already in the structure (internal merge = singulary transformation) and the merger of a lexical item not yet in the structure (external merge = generalized transformation) target the top of the existing tree. Consider in this context the structures in (90)-(92).

```
(61)    X              (62)    X              (63)    X
       / \                    / \                    / \
      Z   A                  β   X                  Z   A
         / \                    / \                    / \
        B   C                  Z   A                  B   C
                                  / \                    / \
                                 B   C                  C   β
```

(61) is the original tree. (62) shows a derivation that obeys the Extension Condition. Here the new element β is merged at the top of the tree. The last derivation, (63), does not obey the Extension Condition because β is merged at the bottom of the tree. Importantly, there is a deep idea behind cyclicity, which again was present in Chomsky's earliest work in the late 1950's. The idea, called the No Tampering Condition in current parlance, seems like a rather natural economy condition. (62) involves no tampering since the old tree in (61) still exists as a subtree of (62), whereas (63) involves tampering with the original structure. That is, it is more economical to expand a structure than to go back and change a structure that has already been built. This becomes particularly clear if parts of the structure are shipped off to the interfaces (e.g., phase by phase as in much recent Minimalist work), where the earlier structure effectively is not available. Were one to tamper with that structure, it would require bringing the structure back into the main structure again, which seems hugely uneconomical.

8. Concluding remarks

The history of generative grammar is not very long. Despite this, considerable progress has been made in our understanding of the human language faculty. Numerous problems and questions remain, but it is interesting to observe that there are certain questions that have remained at the center of the theoretical development since the early beginning. For example, whereas generalized transformations were eliminated in the mid 1960s, they returned again in the Minimalist Program where D-structure was eliminated (though see Uriagereka 2008 for critical discussion). Questions of how structure is generated are still at the forefront of current research. Another major issue, since Ross (1967), is locality (see chapter 18). Since Chomsky (1973), locality issues have occupied a central role in linguistic theorizing. We are still lacking a complete theory of islands, so this is certainly another issue that will be on the front burner for quite some time. Phrase structure has been central since *LSLT*, though the theory of phrase structure has undergone substantial changes over the years.

These are just a few examples of recurrent themes during the brief 60-year history of our field. In this chapter we have in particular emphasized the early period since that is often the period that is not as well known. We believe it is important to know the history of the field in order to fully understand current developments. For example, understanding the change from Government and Binding to the Minimalist Program necessitates a good understanding of the former framework. But in order to understand Government and Binding, it is also necessary to understand the Extended Standard Theory, and in turn also the framework in *LSLT* and *Syntactic Structures* and the one is *Aspects*. We hope that this chapter serves as a useful entry point into this history.

References

Abney, Steven. 1987. The English Noun Phrase in its Sentential Aspect. Doctoral dissertation, MIT.

Bach, Emmon. 1964. *An Introduction to Transformational Grammars*. New York: Holt, Rinehart and Winston.

Baker, Mark C. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Baltin, Mark. 1982. A Landing-site theory of movement rules. *Linguistic Inquiry* 13: 1-38.
den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. *On the Formal Syntax of the Westgermania*, Werner Abraham (ed.), 47-131. Amsterdam: John Benjamins.
Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.
Brame, Michael K. 1976. *Conjectures and Refutations in Syntax and Semantics*. New York: Elsevier.
Bresnan, Joan. 1970. On complementizers: Toward a syntactic theory of complement types. *Foundations of Language* 6: 297-321.
Bresnan, Joan. 1971. Sentence stress and syntactic transformations. *Language* 47: 257-281.
Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Malden: Blackwell.
Brody, Michael. 1995. *Lexico-Logical Form: A Radically Minimalist Theory*. Cambridge, MA: The MIT Press.
Brody, Michael. 2002. On the status of representations and derivations. *Derivation and Explanation in the Minimalist Program*, Samuel David Epstein and T. Daniel Seely (eds.), 19-41. Malden: Blackwell.
Brody, Michael. 2003. *Towards an Elegant Syntax*. London: Routledge.
Chametzky, Robert A. 2000. *Phrase Structure: From GB to Minimalism*. Malden: Blackwell.
Chomsky, Noam. 1955. The Logical Structure of Linguistic Theory. Ms. Harvard University and MIT. [Revised version published in part by Plenum, New York, 1975]
Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions On Information Theory* 2: 113-124.
Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
Chomsky, Noam. 1958/1962. A transformational approach to syntax. *Proceedings of the Third Texas Conference on Problems of Linguistic Analysis in English*, Archibald A. Hill (ed.), 124-158. Austin: University of Texas Press.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.
Chomsky, Noam. 1970a. Remarks on nominalization. *Readings in English Transformational Grammar*, Roderick A. Jacobs and Peter S. Rosenbaum (eds.), 184-221. Waltham, Mass.: Ginn.
Chomsky, Noam. 1970b. Deep structure, surface structure, and semantic interpretation. *Studies in General and Oriental Linguistics Presented to Shirô Hattori on the Occasion of his Sixtieth Birthday*, Roman Jakobson and Shigeo Kawamoto (eds.), 52-91. Tokyo: TEX Company, Ltd.
Chomsky, Noam. 1973. Conditions on transformations. *A Festschrift for Morris Halle*, Stephen Anderson and Paul Kiparsky (eds.), 232-286. New York: Holt, Rinehart and Winston.
Chomsky, Noam. 1977. On *wh*-movement. *Formal Syntax*, Peter Culicover, Thomas Wasow and Adrian Akmajian (eds.), 71-132. New York: Academic Press.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: The MIT Press.

Chomsky, Noam. 1993. A Minimalist program for linguistic theory. *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Kenneth Hale and Samuel J. Keyser (eds.), 1-52. Cambridge, MA: The MIT Press.

Chomsky, Noam. 1995a. Bare phrase structure. *Evolution and Revolution in Linguistic Theory: A Festschrift in Honor of Carlos Otero*, Héctor Campos and Paula Kempchinsky (eds.), 51-109. Washington D.C.: Georgetown University Press.

Chomsky, Noam. 1995b. *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, Roger Martin, David Michaels and Juan Uriagereka (eds.), 89-155. Cambridge, MA: The MIT Press.

Chomsky, Noam. 2001. Derivation by Phase. *Ken Hale: A Life in Language*, Michael Kenstowicz (ed.), 1-52. Cambridge, MA: The MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. *Structures and Beyond: The Cartography of Syntactic Structures*, Adriana Belletti (ed.), 104-131. Oxford: Oxford University Press.

Chomsky, Noam. 2007. Approaching UG from below. *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*, Hans-Martin Gärtner and Uli Sauerland (eds.), 1-30. Berlin: Mouton de Gruyter.

Chomsky, Noam. 2008. On Phases. *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, Carlos Otero, Robert Freidin and María-Luísa Zubizarreta (eds.), 133-166. Cambridge, MA: The MIT Press.

Chomsky, Noam and Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 11: 425-504.

Chomsky, Noam and Howard Lasnik. 1993. The theory of principles and parameters. *Syntax: An International Handbook of Contemporary Research*, Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Venneman (eds.), 506-569. New York: Walter de Gruyter.

Cormack, Annabel. 1999. Without specifiers. *Specifiers: Minimalist Approaches*, David Adger, Susan Pintzuk, Bernadette Plunkett and George Tsoulas (eds.), 46-68. Oxford: Oxford University Press.

Davis, Martin. 1958. *Computability and Unsolvability*. New York: McGraven-Hill Book Company.

Dougherty, Ray C. 1973. A survey of linguistic methods and arguments. *Foundations of Language* 10: 432-490.

Emonds, Joseph. 1976. *A Transformational Approach to English Syntax*. New York: Academic Press.

Epstein, Samuel David. 1999. Un-principled syntax: The derivation of syntactic relations. *Working Minimalism*, Samuel David Epstein and Norbert Hornstein (eds.), 317-345. Cambridge, MA: The MIT Press.

Fiengo, Robert. 1974. Semantic Conditions on Surface Structure. Doctoral dissertation, MIT.

Fiengo, Robert. 1977. On trace theory. *Linguistic Inquiry* 8: 35-62.

Fillmore, Charles J. 1963. The position of embedding transformations in a grammar. *Word* 19: 208-231.

Frank, Robert. 2002. *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, MA: The MIT Press.

Freidin, Robert. 1978. Cyclicity and the theory of grammar. *Linguistic Inquiry* 9: 519-549.
Fukui, Naoki. 2001. Phrase structure. *The Handbook of Contemporary Syntactic Theory*, Mark Baltin and Chris Collins (eds.), 374-406. Malden: Blackwell.
Fukui, Naoki and Margaret Speas. 1986. Specifiers and projection. *MIT Working Papers in Linguistics: Papers in Theoretical Linguistics* 8: 128-172.
Gazdar, Gerrald, Ewald Klein, Geoffrey K. Pullum and Ivan Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.
George, Leland Maurice. 1980. Analogical Generalization in Natural Language Syntax. Doctoral dissertation, MIT.
Halle, Morris. 1959. *The Sound Pattern of Russian: A Linguistic and Acoustical Investigation*. The Hague: Mouton.
Hankamer, Jorge. 1973. Unacceptable ambiguity. *Linguistic Inquiry* 4: 17-68.
Harley, Heidi. 1995. *Subjects, Events and Licensing*. Doctoral dissertation, MIT.
Harman, Gilbert H. 1963. Generative grammars without transformation rules: A defense of phrase structure. *Language* 39: 597-616.
Harris, Zellig. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.
Hoekstra, Eric. 1991. Licensing Conditions on Phrase Structure. Doctoral dissertation, University of Groningen.
Hoekstra, Eric. 1994. Agreement and the nature of specifiers. *Minimalism and Kayne's Antisymmetry Hypothesi*s, C. Jan-Wouter Zwart (ed.), 159-168. Groninger Arbeiten zur Germanistischen Linguistic, Volume 37.
Hornstein, Norbert. 1977. S and X-bar convention. *Linguistic Analysis* 3: 137-176.
Jackendoff, Ray. 1969. Some Rules of Semantic Interpretation for English. Doctoral dissertation, MIT.
Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: The MIT Press.
Jackendoff, Ray. 1977. *X-bar Theory*. Cambridge, MA: The MIT Press.
Jaeggli, Osvaldo. 1980. On Some Phonologically Null Elements in Syntax. Doctoral dissertation, MIT.
Jayaseelan, K. A. 2008. Bare phrase structure and specifier-less syntax. *Biolinguistics* 2: 87-106.
Joshi, Aravind. 1985. How much context-sensitivity is necessary for characterizing structural descriptions? *Natural Language Processing: Theoretical, Computational, and Psychological Perspectives*, David Dowty, Laurie Karttunen and Arnold Zwicky, 206-250. Cambridge: Cambridge University Press.
Joshi, Aravind, Leon S. Levy and Masako Takahashi. 1975. Tree Adjunct Grammar. *Journal of Computer and System Science* 10: 136-163.
Kaplan, Ronald M. and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. *The Mental Representation of Grammatical Relations*, Joan Bresnan (ed.), 173-281. Cambridge, MA: The MIT Press.
Katz, Jerrold J. and Paul Postal. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: The MIT Press.
Kayne, Richard S. 1969. The Transformational Cycle in French Syntax. Doctoral

dissertation, MIT.
Kayne, Richard S. 1975. *French Syntax: The Transformational Cycle*. Cambridge, MA: The MIT Press.
Kayne, Richard S. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris.
Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: The MIT Press.
Koizumi, Masatoshi. 1995. Phrase Structure in Minimalist Syntax. Doctoral dissertation, MIT.
Koster, Jan. 1978. *Locality Principles in Syntax*. Dordrecht: Foris.
Kuroda, Sige-Yuki. 1965. Generative Grammatical Studies in the Japanese Language. Doctoral dissertation, MIT.
Lakoff, George. 1970. Global rules. *Language* 46: 627-639.
Lakoff, George. 1971. On generative semantics. *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, Danny D. Steinberg and Leon A. Jakobovits (eds.), 232-296. Cambridge: Cambridge University Press.
Lasnik, Howard. 1981. Restricting the theory of transformations. *Explanation in Linguistics*, Norbert Hornstein and David Lightfoot (eds.), 152-173 .London: Longmans.
Lasnik, Howard. 2000. *Syntactic Structure Revisited: Contemporary Lectures on Classic Transformational Theory.* Cambridge, MA: The MIT Press.
Lasnik, Howard. 2001. Derivation and representation in modern transformational syntax. *The Handbook of Contemporary Syntactic Theory*, Mark Baltin and Chris Collins (ed.), 62-88. Malden: Blackwell.
Lasnik, Howard and Joseph J. Kupin. 1977. A restrictive theory of transformational grammar. *Theoretical Linguistics* 4: 173-196.
Lasnik, Howard and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15: 235-289.
Lasnik, Howard and Mamoru Saito. 1992. *Move Alpha*. Cambridge, MA: The MIT Press.
Lees, Robert B. 1963. *The Grammar of English Nominalizations*. The Hague: Mouton.
Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
Matthews, G. H. 1965. *Hidatsa Syntax*. The Hague: Mouton.
Muysken, Pieter. 1982. Parametrizing the notion “Head”. *Journal of Linguistic Research* 2: 57-75.
McCawley, James D. 1968. The role of semantics in a grammar. *Universals in Linguistic Theory*, Emmon Bach and Robert T. Harms (eds.), 124-169. New York: Holt, Rinehart and Winston.
Perlmutter, David. 1968. Deep and Surface Constraints in Syntax. Doctoral dissertation, MIT.
Perlmutter, David. 1980. Relational Grammar. *Syntax and Semantics: Current Approaches to Syntax*, Edith A. Moravcsik and Jessica R. Wirth (eds.), 195-229. New York: Academic Press.
Peters, Stanley and R. W. Ritchie. 1973. On the generative power of transformational grammars. *Information Sciences* 6: 49-83.
Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20: 365-424.

Post, Emil. 1944. Recursively enumerable sets of positive integers and their decision problems. *Bulletin AMS* 50. 284-316.

Postal, Paul. 1972. The best theory. *Goals of Linguistic Theory*, Stanley Peters (ed.), 131-170. Englewood Cliffs, NJ: Prentice-Hall.

Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: The MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. *Elements of Grammar*, Liliane Haegeman (ed.), 281-337. Dordrecht: Kluwer.

Ross, John Robert. 1967. Constraints on Variables in Syntax. Doctoral dissertation, MIT. [Published as *Infinite Syntax!* Norwood, N.J.: Ablex, 1986.]

Ross, John Robert. 1969. Guess who? *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society*, Robert I. Binnick, Alice Davison, Georgia M. Green and Jerry L. Morgan (eds.), 252-639. Chicago Linguistics Society, University of Chicago.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris: Payot.

Speas, Margaret J. 1990. *Phrase Structure in Natural Language*. Dordrecht: Kluwer.

Starke, Michal. 2004. On the inexistence of specifiers and the nature of heads. *Structure and Beyond: The Cartography of Syntactic Structures*, Adriana Belletti (ed.), 252-268. Oxford: Oxford University Press.

Stowell, Tim. 1981. Origins of Phrase Structure. Doctoral dissertation, MIT.

Stuurman, Frits. 1985. *Phrase Structure Theory in Generative Grammar*. Dordrecht: Foris.

Travis, Lisa. 1984. Parameters and Effects of Word Order Variation. Doctoral dissertation, MIT.

Uriagereka, Juan. 1999. Multiple Spell-Out. *Working Minimalism*, Samuel David Epstein and Norbert Hornstein (eds.), 251-282. Cambridge, MA: The MIT Press.

Uriagereka, Juan. 2008. *Syntactic Anchors*. Cambridge: Cambridge University Press.

Wasow, Thomas A. 1972. Anaphoric Relations in English. Doctoral dissertation, MIT.

Wells, Rulon S. 1947. Immediate constituents. *Language* 23: 81-117.

Williams, Edwin. 1981. On the notions “lexically related” and “head of a word”. *Linguistic Inquiry* 12: 245-274.