In Norbert Hornstein, Howard Lasnik, Pritty Patel-Grosz and Charles Yang, eds., *Syntactic Structures after 60 Years: The Impact of the Chomskyan Revolution in Linguistics*, de Gruyter Mouton, Berlin, 255-282. (2018)

Transformations in the Quest for a Simpler, more Elegant Theory*

Mamoru Saito

1. Introduction

Syntactic Structures is the work that made transformational grammar widely known, and, to borrow David Lightfoot's expression in the Introduction to the 2002 edition, was "the snowball which began the avalanche of the modern 'cognitive revolution'." The purpose of this chapter is to discuss how transformational rules are motivated in this work and how the relevant discussion and proposals led to the remarkable development of syntactic theory in the subsequent 60 years.

Simply put, the motivation for the transformational rules is that its incorporation makes a grammar much simpler and enables us to capture deeper generalizations. It is however instructive to go over what kinds of simplicity Chomsky had in mind and how the same quest for simplicity and elegance led to later developments in syntactic theory. In order to do this, I discuss the famous analysis of the English auxiliary system in the following section and the passive transformation in Section 3. The proposal of transformations in *Syntactic Structures* started intense investigations on their properties, which in turn made it possible to raise the fundamental question why Language has transformations. In Section 4, I consider the recent developments on this question in the Minimalist framework.

Before I start the discussion of transformations, let me briefly comment on 'simplicity of the overall system' as a motivation because this itself marked a sharp departure from the dominant view on linguistic theory at the time. In Chapter 6, titled "On the Goals of Linguistic Theory," Chomsky lays out the methodological foundation for linguistic research. He first argues against the position widely held in American structuralism that it is among the goals of linguistic theory to provide a discovery procedure for grammars, a manual of mechanical method to construct a grammar on the basis of a given corpus. His alternative is stated as follows:

> The point of view adopted here is that it is unreasonable to demand of linguistic theory that it provide anything more than a practical evaluation procedure for grammars. (p.52)

* I would like to thank Norbert Hornstein for his helpful feedback on the initial version of this manuscript.

An evaluation procedure is a way to choose among proposed grammars. And the criterion for the evaluation, aside from adequacy in description, is nothing but simplicity. This, Chomsky states, necessitates investigation into the notion of simplicity itself. At the same time, he proceeds with syntactic analysis with the following statement:

> Nevertheless, it should be fairly clear that under any reasonable definition of "simplicity of grammar," most of the decisions about relative complexity that we reach below will stand. (p.55)

Transformations, then, are motivated because they yield simpler, or more elegant, grammars.

Chomsky emphasizes here that the simplicity at issue is not simplicity at a particular level or within a particular component of grammar but is overall simplicity. In this context, he also argues against another idea that is closely associated with a discovery procedure, that is, that linguistic research should proceed in a bottom-up fashion, from phonemic analysis to morphology and then to syntax. Chomsky notes that "the higher levels of linguistic description depend on results obtained at the lower levels" but "the converse is true" as well. (p.59) This is amply demonstrated throughout the discussion in *Syntactic Structures*. Then, we simply try to formulate the general theory of syntax and the grammars of particular languages as precisely and as elegantly as possible, and continue to develop them on the basis of empirical and conceptual considerations. This is taken for granted in the current research in generative grammar, which makes no methodological distinction between linguistics and (other) natural sciences. But it is worth noting that it is this approach to linguistics that led to the proposal of transformational rules over 60 years ago.

2. Inadequacy of Phrase Structure Rules for Discontinuous Elements

In this section, I briefly go over the analysis of the English Auxiliary system in *Syntactic Structures* and discuss the model of syntax it leads to.

2.1. English Auxiliary System

After Chomsky introduces phrase structure rules of the following kind in Chapter 4 to capture constituent analysis, he points out their inadequacies in Chapter 5:

(1) a. *Sentence* → *NP + VP*
b. *VP* → *Verb + (NP)*
c. *NP* → *(Art) + N*
d. *Art* → *the*
e. *N* → *man, ball, book,* etc.
f. *Verb* → *hit, took,* etc.

One case concerns the auxiliary system. (2) illustrates the point.

(2) The man had been reading the book.

Here, the perfect is expressed by *have* and the affix *en* on the following element, and the progressive by *be* and the affix *ing* on the following verb. In other words, the perfect and the progressive are expressed by discontinuous elements. Phrase structure rules fail to capture elements of this kind.

Chomsky then goes on to propose the celebrated transformational analysis. He first adds the phrase structure rules in (3) to (1a-e).

(3) a. *Verb* → *Aux + V*
b. *Aux* → *Tense + (Modal) + (have + en) + (be + ing)*[1]
c. *V* → *hit, take, walk, read,* etc.
d. *Tense* → *past, present*
e. *Modal* → *will, can, may, shall, must*

Note that *have* and *be* are contiguous with their associated affixes in (3b). Then, he introduces the rules in (4) to derive the surface form.

(4) a. Let *Af* stand for any of the affixes *past, present, en, ing*. Let *v* stand for any *Modal* or *V*, or *have* or *be* (i.e., for any non-affix in the phrase *Verb*.) Then:
Af + *v* → *v* + *Af* #, where # is interpreted as word boundary.
b. Replace + by # except in the context *v* – *Af*. Insert # initially and finally.

(4a), which later came to be called 'affix hopping' and more recently 'phonological merger',

[1] In *Syntactic Structures*, *C* is employed instead of *Tense* and it expands to *past*, *s* (for singular subject) or ∅ (for plural subject). I use the more familiar *Tense* here and accordingly in (3d).

moves affixes to the end of the following elements. (4b) specifies the word boundaries.

(5) illustrates how (2) is derived by the rules in (1), (3) and (4).

(5) a. *the + man + past + have + en + be + ing + read + the + book*
b. *the + man + have + past # be + en # read + ing # the + book*
c. *# the # man # have + past # be + ing # read + ing # the # book #*

The phrase structure rules generate the string in (5a). (4b) applies to this string and places *past* after *have*, *en* after *be*, and *ing* after *read* as in (5b). (4c) inserts word boundaries except before the suffixes to yield (5c). Finally, it is assumed that morphophonemic rules of the following form apply to (5c):

(6) *walk* → /wɔk/, *take + past* → /tʊk/, *have + past* → /həd/, *be + en* → /bɪn/, …

What matters here is the simplicity and the elegance of this analysis. Chomsky never claimed that phrase structure rules cannot accommodate the English auxiliary system. The rules in (7), for example, can be employed instead of (3b) and (4a) to generate the grammatical patterns.

(7) a. *Verb → V + Tense*
b. *Verb → Modal + Tense + V*
c. *Verb → have + Tense + V + en*
d. *Verb → be + Tense + V + ing*
e. *Verb → Modal + Tense + have + V + en*
f. *Verb → Modal + Tense + be + V + ing*
g. *Verb → have + Tense + be + en + V + ing*
h. *Verb → Modal + Tense + have + be + en + V + ing*

But this is clearly more complex than the analysis that employs (4a), and misses the generalization that the perfect is expressed by two discontinuous elements *have* and *en* and the progressive by *be* and *ing*.

The analysis outlined above leads to the model of grammar quoted in (8) from p.46 of *Syntactic Structures*.

(8) Σ: Sentences:

F: $X_1 \rightarrow Y_1$
.
.
$X_n \rightarrow Y_n$ } Phrase structure

T_1
.
.
T_n } Transformational structure

$Z_1 \rightarrow W_1$
.
.
$Z_n \rightarrow W_n$ } Morphophonemics

Phrase structure rules generate strings with structures that constitute inputs to transformational rules. Then, transformations yield strings that morphophonemic rules apply to. This raises the question whether the outputs of phrase structure rules and transformations encode specific information and if they do, what sorts of information they express. The pursuit of this question led to the postulation of D-structure, S-structure and PF in the Standard Theory of the 1960's and the Extended Standard Theory (EST) later.

2.2. The Generality of the Affix Hopping Analysis

The beauty of the affix hopping rule in (4a) is that it applies quite generally to any structure regardless of how the structure is created, and further, leads to an explanation of an otherwise mysterious phenomenon, namely, the appearance of the "dummy verb" *do* in some contexts. Let us consider a few cases from *Syntactic Structures*.

Chomsky observes that a sentence can be analyzed as having three parts, 1 – 2 – 3, as in (9), and elements of negation, *not* and *n't*, occur after the second.

(9) a. *NP – Tense – V ...*
b. *NP – Tense + Modal – ...*
c. *NP – Tense + have – ...*
b. *NP – Tense + be – ...*

Thus, *not* appears after *have* and *be* in (10a) and (10b) respectively.

(10) a. The man had not taken the book.
b. The man was not taking the book.

The derivation of (10a) is illustrated in (11).

(11) a. *the + man + past + have + en + take + the + book*
b. *the + man + past + have + not + en + take + the + book*
c. *the + man + have + past # not + take + en # the + book*
d. *# the # man # have + past # not # take + en # the # book #*

The phrase structure rules generate the string in (11a), and then, *not* is inserted after *have* in (11b).[2] The affix hopping transformation in (4a) applies as in (11c), placing *past* after *have* and *en* after *take*. The example shows that (4a) serves to generate negative sentences as well as affirmative ones.

The interesting case is (9a), instantiated by (12).

(12) The man did not take the book.

The derivation of this example is as in (13).

(13) a. *the + man + past + take + the + book*
b. *the + man + past + not + take + the + book*
c. *# the # man # past # not # take # the # book #*
d. *# the # man # do + past # not # take # the # book #*

The insertion of *not* after *Tense* blocks the application of (4a) to *Tense + V* as (13b) shows. Hence, *take* does not appear in past tense and *past* as an affix is stranded after the insertion of word boundaries in (13c). Chomsky, then, proposes the rule in (14), now known as '*do*-support', in order to save the stranded tense affix.

(14) *# Af* → *# do + Af*

[2] It is thus proposed in *Syntactic Structures* that negative elements are inserted into the structure by a transformation. But this is not important for the point to be made here, that is, the generality of the application of the affix hopping rule in (4a).

This rule inserts *do* before a stranded affix as in (13d) and yields (12) correctly. Note that this analysis is possible because the phrase structure rule (3b) treats *Tense* and *V* as independent elements and (4a) serves to combine them when they are adjacent. If *V* + *Tense* enters the structure as a single element, there is no straightforward way to account for the appearance of *do* in (12).

Chomsky goes on to show that basically the same analysis applies to other independent constructions. One of them is *yes/no* questions, exemplified in (15).

(15) a. Had the man taken the book?
b. Was the man taking the book?
c. Did the man take the book?

Yes/no questions are formed by the question transformation that fronts the second element in (9) to the sentence-initial position. And interestingly, *do* appears in (15c) just as in (12). The derivation of the example is shown in (16).

(16) a. *the* + *man* + *past* + *take* + *the* + *book*
b. *past* + *the* + *man* + *take* + *the* + *book*
c. # *past* # *the* # *man* # *take* # *the* # *book* #
d. # *do* + *past* # *the* # *man* # *take* # *the* # *book* #

The fronting of *past* by the question transformation in (16b) blocks the application of (4a) and produces a stranded *Tense* as in (16c). This triggers the *do*-support rule in (14) and yields the form in (15c). Thus, the forms of questions are explained by exactly the same set of rules as negative sentences.[3]

It is mentioned in *Syntactic Structures* that the apparent asymmetry in (17) follows as a consequence of the analysis.

---

[3] Chomsky points out that the same set of rules accounts for examples of conjunction with *so* as in (i).

(i) a. John has arrived and so have I.
b. John arrives and so do I.

It extends to examples of VP-ellipsis and VP-preposing as well, as it is now standardly taught in introductory syntax courses.

(17) a. What did the man take?
b. Who took the book?

As (17a) indicates, a wh-question is generated with further fronting of a wh-word (or phrase).[4] However, *do* does not show up and there is apparently no word-order change in (17b) with a wh-word in the subject position. It seems then that neither the question rule nor the wh-fronting rule applies in this case. Are wh-questions generated in two distinct ways depending on the location of the wh-word? Chomsky's answer is negative. The derivation of (17b) with the question transformation and wh-fronting is shown in (18).

(18) a. *who + past + take + the + book*
b. *past + who + take + the + book*
c. *who + past + take + the + book*
d. *who + take + past # the + book*
e. *# who # take + past # the # book #*

The question transformation fronts *past* to the sentence-initial position as in (18b). At this point, the subject intervenes between *past* and *take*. However, the wh-fronting rule fronts the wh subject as in (18c), making *past* and *take* adjacent again. Then, the affix hopping rule (4a) can apply and form *take + past*. Thus, the question transformation and the wh-question rule apply uniformly regardless of the location of the wh-word. The analysis is clearly simpler than having two sets of distinct phrase structure rules for subject and non-subject wh-questions.

In the discussion above on the affix hopping rule (4a), I introduced a few more transformations from *Syntactic Structure*. Each of them can be motivated on the basis of simplicity. For example, one may assume the phrase structure rules in (19) and (20) in place of the question transformation.

(19) a. *Sentence → Do + Tense + NP + VP_1*
b. *Sentence → Modal + Tense + NP + VP_2*

---

[4] More precisely, the transformation is defined in Footnote 2 on p.69 as a rule that adds *wh* to pronouns and fronts *wh + pronoun*. Then, morphophonemic rules of the following form apply:

(i) *wh + he/she* → /huw/, *wh + it* → /wat/

I assume for the sake of simplicity that phrase structure rules directly generate strings with wh-words.

c. *Sentence* → *have* + *Tense* + *NP* + VP_3
d. *Sentence* → *be* + *Tense* + *NP* + VP_4

(20) a. VP_1 → $Verb_1$ + *(NP)*
b. $Verb_1$ → *V*
c. VP_2 → $Verb_2$ + *(NP)*
d. $Verb_2$ → *have* + *be* + *en* + *V* + *ing*
e. $Verb_2$ → *have* + *V* + *en*
f. $Verb_2$ → *be* + *V* + *ing*
g. $Verb_2$ → *V*
h. VP_3 → $Verb_3$ + *(NP)*
i. $Verb_3$ → *be* + *en* + *V* + *ing*
j. $Verb_3$ → *V* + *en*
k. VP_4 → $Verb_4$ + *(NP)*
l. $Verb_4$ → *V* + *ing*

It should be obvious that this is far more complex than the question transformation and that it misses generalizations on the formation of questions and the placement of affixes. According to the analysis in *Syntactic Structures*, questions are formed by fronting the second element in (9) and the rest of the derivation is automatically taken care of by the independently motivated affix hopping rule and *do*-support. On the other hand, it is not clear that the complex set of rules in (19) and (20) expresses any generalization. For example, given these rules, it is simply an accident that what appears before the subject NP in (19) is missing from the *VP* in the same rule. One can in fact easily restate the rules so that this generalization does not obtain.

2.3. Refinements of the Analysis

As argued and demonstrated in Lasnik (1995), the essential part of the analysis introduced in this section still remains a viable possibility even today, 60 years after it was proposed. It was in fact refined in the details over the years.[5]

Recall that the formulations of negative placement and question transformations are based on the tri-part analysis of sentences as in (9), repeated below in (21).

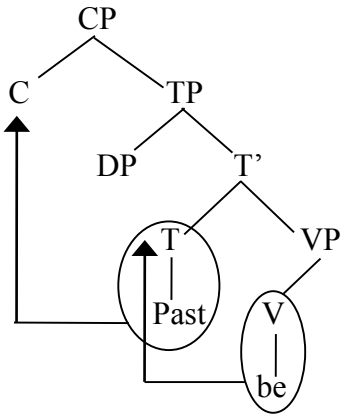
[5] See, in particular, Lasnik (1981) and the references cited there.

(21) a. *NP – Tense – V ...*
b. *NP – Tense + Modal – ...*
c. *NP – Tense + have – ...*
d. *NP – Tense + be – ...*

This indicates that *Tense* forms a unit with the following *Modal*, *have* or *be*, but not with a main verb *V*. This is captured by the generation of *Modal* under *Tense* and the head-movement of *have* and *be* to *Tense*. Then, the elements of negation, *not* and *n't*, occur right after *Tense*, and it is *Tense* that is fronted in question formation. The derivation of (15b), repeated in (22), in more modern terms is shown in (23).

(22) Was the man taking the book?

(23)



Then, *Tense* combines with *have* and *be* by head movement, and with the main verb by affix hopping.

There is a remark in *Syntactic Structures* on why the forms in (24) are both attested in dialects whereas (25b) is excluded.

(24) a. Have you the book?
b. Do you have the book?

(25) a. Are you comfortable?
b. *Do you be comfortable?

It is stated that (24a) and (24b) are both permissible because the main verb *have* can be analyzed as *have* in (21c) or as *V* in (21a). The proposal for (25) is that the main verb *be* does not belong to the class of verbs and hence can only be analyzed as *be* in (21d).

The head movement analysis captures the facts more straightforwardly. *Be*, whether it is an auxiliary verb or a main verb undergoes head movement to *Tense* when it is the first element among auxiliaries and verbs. Whether the main verb *have* moves to *Tense* depends on the dialect. Emonds (1978) extends the analysis to French where all main verbs are raised to *Tense* in appropriate contexts. Thus, contrasts such as that in (26) obtain between French and English.

(26) a. Jean (n') aime pas Marie.
(*ne*) love Neg
b. John does not love Mary.

The main verb precedes *pas* in (26a) as it raises to *Tense* whereas it follows *not* in (26b).

3. Phrasal Movements and Selection

There are very important discussions on passive in *Syntactic Structures*, which pave the way to the postulation of D-structure in later works. In this section, I go over the presented arguments for the passive transformation and discuss their consequences.

3.1. The Passive Transformation

The passive transformation is proposed in the following form in Chapter 4:

(27) If S_1 is a grammatical sentence of the form $NP_1 - Aux - V - NP_2$,
then the corresponding string of the form $NP_2 - Aux + be + en - V - by + NP_1$
is also a grammatical sentence.

This rule derives (28b) as in (29).

(28) a. Sincerity frightens John.
b. John is frightened by sincerity.

(29) a. *sincerity + present + frighten + John*
b. *John + present + be + en + frighten + by + sincerity*
c. *John + be + present # frighten + en # by + sincerity*
d. *# John # be + present # frighten + en # by # sincerity #*

The phrase structure rules generate (29a), which underlies the active sentence (28a). The passive transformation applies to this string, yielding (29b). The rule exchanges the positions of the subject and the object, adds *by* before the postposed subject, and adds *be + en* after *Aux*. The affix hopping rule applies to the output of the passive rule as in (29c), and finally (29d) is produced as the input to morphophonemics.

Chomsky considers the possibility to introduce *be + en* as part of *Aux* in the phrase structure rule as in (30).

(30) *Aux → Tense + (Modal) + (have + en) + (be + ing) + (be + en)* (cf. (3b))

However, he rejects the possibility as heavy restriction must be placed on the selection of *be + en*. The following *V* must be transitive and at the same time the *V* cannot be followed by *NP*. The statement of these complicates the grammar.

It is worth noting here that the now standard argument on the basis of selectional relations is already presented in *Syntactic Structures* for the passive rule. The comparison of (28a) and (31a) shows that *frighten* selects an animate object.

(31) a. # John frightens sincerity.
b. # Sincerity is frightened by John.

And the ungrammaticality of (31b), as opposed to (28b), indicates that the same selectional requirement is imposed on the subject instead of the object in the passive counterpart. In short, the selectional requirements on the object and the subject are reversed in passives. If passive sentences are directly generated by phrase structure rules, the selectional requirements must be stated separately for active and passive sentences, which would be an inelegant reduplication. With the passive transformation, the same requirements can be imposed on the output of phrase structure rules uniformly for both active and passive sentences.

It is sometimes stated that the passive transformation is motivated because an active-passive

pair are synonymous. It is also often assumed that the active sentence is the "basic form" because it directly represents the predicate-argument structure. Neither of these is assumed in *Syntactic Structures*. In Chapter 7, Chomsky considers the possibility that the passive rule only introduces *be* + *en* and *by*, and does not exchange the positions of the subject and the object (or more precisely the NP that immediately follows *V*). He notes first that this would complicate the statement of selectional requirements because the rule derives (31b) from what underlies (28a). Then, he goes on to present another similar argument against this option on the basis of the contrast between (32) and (33).

(32) a. All the people in the lab consider John a fool.
b. John is considered a fool by all the people in the lab.

(33) a. * John considers all the people in the lab a fool.
b. * All the people in the lab are considered a fool by John.

Chomsky analyzes *consider* + *a fool* in (32a) as a complex *V* that selects a singular object. The object *John* is then placed before *a fool* by a transformation. If the passive transformation is formulated as in (27) and makes NP_2 the new subject, (32b) is correctly predicted to be grammatical. The selectional requirement of *consider* + *a fool* is met by the singular NP, *John*, before the transformation applies. The rule also correctly predicts (33b) to be ungrammatical. *All the people in the lab* originates in the object position and it contradicts the selectional requirement of the complex *V* in that position. Thus, (33b) is ruled out in exactly the same way as (33a).

On the other hand, if the passive rule preserves the positions of the subject and the object, (33b) would be derived from (32a). A stipulation would then be required to exclude the example. Similarly, it would have to be explained why (32b) is grammatical despite the fact that it is derived from a string that underlies the ungrammatical (33a). Then, the overall simplicity is attained with the formulation of passive that exchanges the positions of NP_1 and NP_2 as in (27). Note that this argument stands independently of the complex *V* analysis of *consider* + *a fool*. Suppose that *John* + *a fool* in (32a) constitutes a small clause with the predicate *a fool* and the subject *John* as proposed by Stowell (1981). Then, the predicate *a fool* selects a singular subject. The *NP*, *John*, satisfies this requirement in the example. With the passive transformation in (27), exactly the same analysis can be given to (32b). The example is grammatical because the subject, *John*, originates as the subject of the small clause. (33b), on the other hand, is not because the subject, *all the people in the lab*, fails to satisfy the selectional requirement of the predicate, *a fool*, in its initial position.

Chomsky also argues on the basis of simplicity that a passive sentence should be derived from what underlies its active counterpart and not vice versa. The specific argument given in this context is based on the examples in (34).

(34) a. The wine was drunk by the guests.
b. John was drunk by midnight.

(34a) is a passive sentence whereas *drunk* in (34b) is an adjective, as the following contrast attests:

(35) a. * The wine was very drunk by the guests.
b. John was very drunk by midnight.

If the "active transformation" applies to what underlies (34a) to generate (36a), then it would incorrectly generate (36b) as well.

(36) a. The guests drank wine.
b. * Midnight drank John.

Stronger arguments can be built also on the basis of simplicity and elegance. For example, if active transitive sentences are produced only by "active transformations," then there should be two distinct ways to generate active sentences. When *V* is not immediately followed *NP*, the sentence is generated by phrase structure rules. On the other hand, a sentence with a *V + NP* sequence is produced by the "active transformation." This clearly misses important generalizations that apply to all active sentences. Suppose instead that active transitive sentences can be generated either by phrase structure rules or by the "active transformation." Then, the transformation would just be redundant. In any case, what is noteworthy here is that Chomsky considers all possibilities without preconception, including the derivation of active from passive, and argues for a set of specific proposals on the basis of the simplicity of syntactic analysis and the overall system. And as will be discussed in the following subsection, the set of proposals in *Syntactic Structures* lays the foundation for research in syntactic theory for the next 40 years.

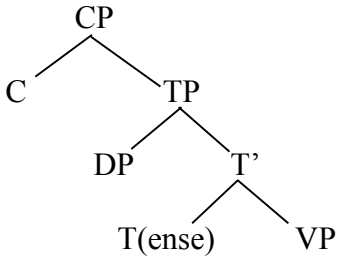
3.2. The Emergence of a Model for Syntax

Chomsky argued for transformations in pursuit of the simplest and most elegant theory, and arrived at the model of derivation in (8), which can be restated as in (37).

(37) phrase structure rules → transformations → morphophonemic rules

This is only a step away from the Standard theory of Chomsky (1965) with D-structure as the output of phrase structure rules (and lexical insertion) and S-structure as the output of the transformational component, and from the Extended Standard Theory pursued, for example, in Chomsky (1981), with the added level of representation, Logical Form. Particularly important in this context is the proposal that selectional relations are represented at the output of phrase structure rules before transformations apply. This leads to the postulation of D-structure as a pure representation of selectional or thematic relations. The pursuit of the simplest theory of the constituent structure, described initially by the phrase structure rules, culminated into X' principles applying at D-structure. Accordingly, the clause structure was reanalyzed as in (38).[6]

(38)



Syntactic Structures also suggests directions for later research on the transformational component and its output, S-structure. Relevant here is its discussions on rule ordering and optional vs. obligatory transformations. I briefly go over them in this subsection.

Rule ordering is first introduced with respect to the affixing hopping rule (4) and *do*-support in (14), repeated below as (39) and (40) respectively.

(39) a. Let *Af* stand for any of the affixes *past, present, en, ing*. Let *v* stand for any *Modal* or *V*, or *have* or *be* (i.e., for any non-affix in the phrase *Verb*.) Then:
Af + *v* → *v* + *Af* #, where # is interpreted as word boundary.

b. Replace + by # except in the context *v* – *Af*. Insert # initially and finally.

(40) # *Af* → # *do* + *Af*

Let us consider (41a), which underlies (41b).

---

[6] See, for example, Chomsky (1970, 1986a), Stowell (1981), and the references cited there for this development on clause structure.

(41) a. *the* + *man* + *past* + *take* + *the* + *book*
b. The man took the book.

If (39b) is applied first to (41a) as in (42a), then the context is set for the application of *do*-support as in (42b).

(42) a. # *the* # *man* # *past* # *take* # *the* # *book* #
b. # *the* # *man* # *do* + *past* # *take* # *the* # *book* #

This generates (43) incorrectly in non-emphatic context.

(43) * The man did take the book.

Then, (39a) must apply before (39b) and (40), and (40) must be construed as a "last resort" rule, an assumption that is maintained even today.

More serious cases of rule ordering arise when the interaction of (39)-(40) and the wh-fronting rule is considered. (44) is a relevant case.

(44) Who took the book?

(45a) is the input to the transformational component.

(45) a. *who* + *past* + *take* + *the* + *book*
b. *past* + *who* + *take* + *the* + *book*
c. # *past* # *who* # *take* # *the* # *book* #
d. # *do* + *past* # *who* # *take* # *the* # *book* #
e. # *who* # *do* + *past* # *take* # *the* # *book* #

The question rule fronts *past* as in (45b). The application of (39)-(40) to this string yields (45d). Finally, the wh-fronting rule places the wh-word in the sentence-initial position as in (45e). This results in (46), which is again illicit in non-emphatic context.

(46) * Who did take the book?

The problem here is that (39)-(40) apply before the wh-fronting rule. If wh-fronting is ordered

before (39)-(40), only the grammatical form is generated as in (47).

(47) a. *who* + *past* + *take* + *the* + *book*
b. *past* + *who* + *take* + *the* + *book*
c. *who* + *past* + *take* + *the* + *book*
d. *who* + *take* + *past* # *the* + *book*
e. # *who* # *take* + *past* # *the* # *book* #

Then, (39)-(40) are ordered after the other transformations. This was taken later as evidence that these rules map the output of the transformational component to the morphophonemic or phonological component. The revised model of derivation in (48) obtains.

(48) phrase structure rules → D-structure → transformations
→ affix-hopping/*do*-support → phonological component

In addition to rule ordering, there is what appears to be an ordering paradox in the analysis in *Syntactic Structures*. *Tense*, which is called *C* in *Syntactic Structures*, is expanded by the following rule:

(49) $C \rightarrow \left\{ \begin{array}{l} S \text{ in the context of } NP_{sing} \_ \\ \emptyset \text{ in the context of } NP_{pl} \_ \\ past \end{array} \right\}$

This looks like a phrase structure rule although it is context-sensitive. However, as Chomsky points out, it must apply after the passive transformation as *present* agrees with the surface subject. Then, (49) is included among the transformational rules in *Syntactic Structures*. But, aside from the fact that it has a peculiar format as a transformation, it leads to further rule ordering. For example, it not only has to apply after passive but also has to apply before the question transformation in examples like (50).

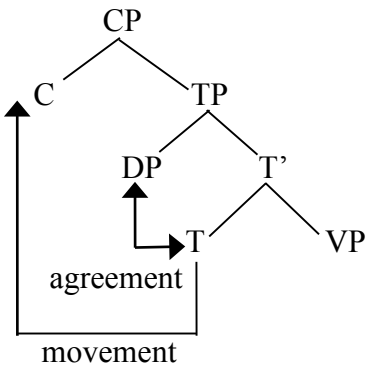
(50) What do the men eat?

In this example, *present* agrees with *the men*, not with *what*.

The ordering of "agreement" with passive is avoided in later works such as Chomsky (1981) by making agreement apply or be checked after the transformational component, that is, at S-

structure. The copy (or trace) theory of movement accommodates (50) as illustrated below.

(51)

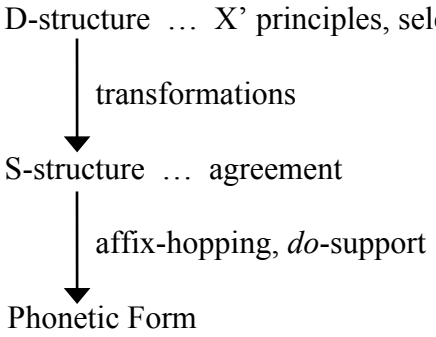


The movement of *T(ense)* to *C* leaves a copy behind, and the copy participates in agreement with the subject. The copy theory, in turn, suggests that the passive transformation does not involve rightward movement of the subject NP and simply moves the object NP to a "vacant" subject position. If the subject NP moves rightward, its copy should block the movement of the object NP. This simplification of passive is supported by the optionality of *by*-phrase as shown in (52).[7]

(52) The book was stolen (by a thief).

The considerations above have narrowed the range of transformations and have led to the following model:

(53)



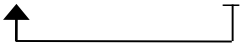
The phrase structure rules specified the possible constituent structures and raised the question why those structures and only those structures are possible. This led to the proposal of X' principles, which resulted in radical simplification of syntactic theory and particular grammars. Similarly, the transformational rules described how movement takes place and raised the question why movement

[7] The treatment of *by*-phrase under this approach poses interesting questions and has been discussed over the years. See, for example, Jaeggli (1986), Baker, Johnson and Roberts (1989), and Collins (2005).

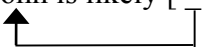
applies in the way it does. This was a major research topic in the 1970's that led to the LGB theory of Chomsky (1981).

Important in this research topic was the distinction between obligatory and optional applications of transformations raised in *Syntactic Structures*. There, the passive transformation which derives a passive sentence from what underlies its active counterpart is mentioned as a typical optional transformation. If it applies, a passive sentence is generated, and if it does not, the derivation proceeds to generate an active sentence. On the other hand, the affix hopping rule was included among the obligatory transformations. The later research on transformations focused on NP movement to subject position as in passive and operator movement to Spec, CP as in wh-question movement, and the question on the obligatory vs. optional applications of transformations took a somewhat different shape. For example, movement to the subject position is apparently obligatory in the example of passive in (54) and the example of raising in (55).

(54) a. * (There) was stolen a book.
b. A book was stolen _ .



(55) a. * (It) is likely [John to succeed].
b. John is likely [_ to succeed].



The question was why the movements in (54) and (55) are obligatory.

An answer was proposed in Chomsky and Lasnik (1977), which showed that independently of movement, the distribution of NPs in English is limited to the four positions in (56).

(56) a. the position following a transitive verb
b. the position following a preposition
c. the subject position of a tensed clause
d. the subject position of a noun phrase

(54a) and (55a) are ungrammatical simply because *a book* and *John* appear in illicit positions in these examples. As is well known, this kind of analysis led to the unification of all movement transformations as *Move-α*, which states, "Move anything anywhere optionally." *Move-α* may but need not apply to *John* in (55a). If it does not, the sentence is ruled out as *John* appears in an illicit

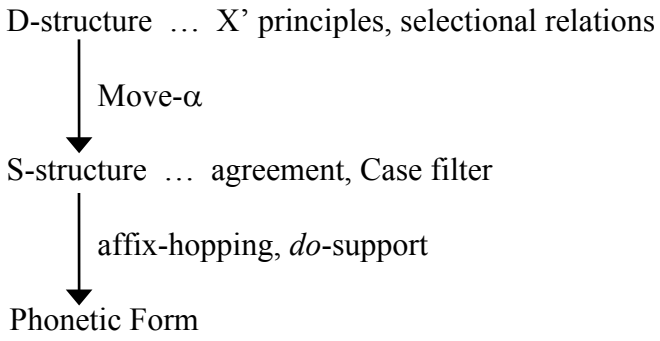
position. If it applies to *John* and move it to the end of the sentence, the result is still ungrammatical because the NP ends up in a position that is not included in (56). Finally, if *Move-α* yields (55b), locating *John* in the matrix subject position, then the sentence is grammatical because the NP is in the subject position of a tensed clause, the position specified in (56c).

It was Jean-Roger Vergnaud's Case theory that provided an explanation for the distribution of NPs in (56). He proposed that nouns need to be specified for Case to assume proper phonetic forms, as can be witnessed in the paradigms of pronouns as in *he – him – his*, and (56) lists the positions where nouns can receive Case. For example, (56a) and (56c) are the positions for accusative and nominative respectively. Chomsky (1981) proposed the Case filter in (57) to formally accommodate the idea.

(56) * NP if NP has phonetic content and no Case.

The Case filter applies at S-structure as indicated in the revised model in (57).

(57)



(57) shows only part of the model proposed in Chomsky (1981), but its significance should be clear. Syntactic theory consists of levels of representations, in particular, D-structure and S-structure, and the principles that define them. Transformations are now reduced to a single principle on the relation between the two levels, *Move-α.* As often noted, this revolutionalized syntactic theory as it conceives of grammar as a set of principles (and parameters) instead of a system of rules. Yet, it is a direct descendent of the model of *Syntactic Structures* in many ways. Notably, the very same quest for simplicity and elegance led to this model. The transformational component is radically simplified and the addition of the Case filter not only aids the simplification but broadens the empirical coverage of the overall theory. And the core part of the model retains its shape. First, the basic constituent structure is specified, reflecting the selectional relations. Then, transformations apply to this structure and produce the input to morphophonemic (or morphophonological) operations.

4. A Minimalist Perspective on Transformations

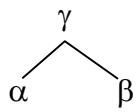
It was seen in the preceding sections that the LGB theory of Chomsky (1981) can be viewed as a direct descendant of the model proposed in *Syntactic Structures*. In both, there is a sharp distinction between what generates the basic phrase structure, phrase structure rules in the case of *Syntactic Structures*, and transformations. They are distinct in form and the latter applies to the output of the former. An effort to eliminate this distinction was initiated in the Minimalist research, in particular, in Chomsky (1994). Although this led to departure from the conception of transformations in *Syntactic Structures*, I discuss it in this section because it addresses a fundamental question that originates in the work, namely, what transformations are, and brings us right to the current research. Section 4.1 concerns the last resort nature of phrasal movement transformations and Section 4.2 discusses the issue of movement of NPs into selected positions.

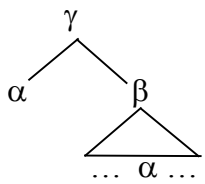
4.1. Merge and the Last Resort Nature of Phrasal Movement

The Minimalist model proposed in Chomsky (1993) eliminates D-structure and S-structure. Pursuing this model, Chomsky (1994) proposes that phrase structure is constructed by *Merge*, the minimal operation required for Language. It takes two objects α and β, and forms a constituent as in (58).

(58) $\gamma = \{\alpha, \beta\}$

When α and β are independent objects as in (59a), the operation is called external Merge.

(59) a. 

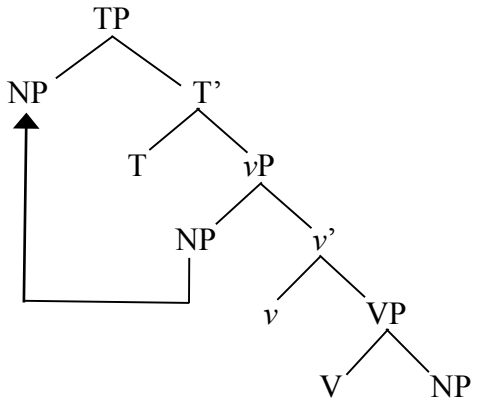
b. 

On other hand, when α is contained within β, and *Merge* applies α and β as in (59b), it is called internal Merge. Both are instances of the minimal operation *Merge*, and phrasal movement is nothing but internal Merge.

The derivation of the simple example in (60) is illustrated in (61).

(60) The man took the book.

(61)



The structure assumes the predicate-internal subject hypothesis as formulated in Chomsky (1995), and consequently, the subject NP is initially merged at the specifier position of *v*P.[8] The structure is straightforwardly built with *Merge*, including internal Merge of the subject NP to the surface subject position.

The conception of phrasal movement as an instance of *Merge* eliminates its special status, and answers the question why phrasal movement exists. Language has phrasal movement just because it has the minimally required operation, *Merge*. However, a few more steps were required to make this hypothesis truly feasible. Recall that movement transformations were unified as a single principle, *Move-α*, which allows movement of anything anywhere. This overgenerates vastly as the examples in (62) and (63) show.

(62) a. The man ate nothing.
b. * Nothing, the man ate _ .

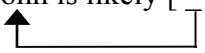
(63) a. The man did not eat anything.
b. * Anything, the man did not eat _ .

The objects in (62a) and (63a) are merged with TP in (62a) and (63b) respectively, and the results are ungrammatical. A proposal to avoid this problem is made in Chomsky (1986b). He notes that

[8] See, for example, Koopman and Sportiche (1991) for arguments for the predicate-internal subject hypothesis.

the movement in (64), for example, applies with "a purpose."

(64) John is likely [_ to succeed].



If *John* stayed in the original position, it cannot receive Case. Hence, it must move to the matrix subject position for Case reasons. Chomsky then proposes that movement, or internal Merge, applies as a last resort, that is, only when it is necessary for the moved element. This is incorporated into the definition of movement in Chomsky (1995). But then, internal Merge is distinguished from external Merge as there is no such restriction on the latter.

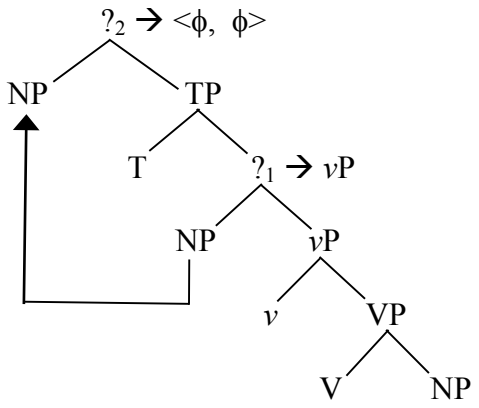
A solution to this problem, as far as I know, was first provided in Chomsky (2013). There he proposes that *Merge* must accompany a labeling algorithm that specifies the nature of the newly formed object. When a verbal element and a nominal element are combined, for example, the interpretive component must be informed whether the formed object is verbal (VP) or nominal (NP). Chomsky first notes that the three cases of *Merge* in (65) must be considered in this respect.

(65) a. $\gamma = \{H, \alpha P\}$
b. $\gamma = \{\alpha P, \beta P\}$
c. $\gamma = \{H_1, H_2\}$

A head and a phrase are merged in (65a). Chomsky states that the head determines the label of γ in this case as search into γ immediately yields a unique head. On the other hand, (65b) and (65c), the mergers of two phrases and two heads respectively, are problematic because the labels cannot be determined straightforwardly in this way. Chomsky notes, however, that (65b) occurs in actual derivations and make two proposals for this case.

Let us consider the derivation of (60) again, repeated in a somewhat different form in (66).

(66)



First, the verb *take* merges with the object *the book*. This is unproblematic as it instantiates (65a), the merger of a head and a phrase. Then, *v* merges with VP, which is also straightforward. But then, a problem arises when the subject NP *the man* merges with *v*P. Chomsky notes that the subject NP internally merges with TP after T is introduced into the structure. Then, $?_1$ only dominates a copy of this NP and does not contain it in full. He proposes that this disqualifies the NP as a label provider for $?_1$ and consequently *v*P does the work. When the subject internally merges with TP, another configuration of (65b) is created. In this case, T and NP share the same ϕ-features (person, number and gender features) because of agreement. Chomsky proposes that this enables $?_2$ to be labeled <ϕ, ϕ>. This captures the fact that movement always terminates with the internal merge of two phrases that share some features.

Chomsky extends the analysis to wh-question movement, as illustrated in (67).

(67) [$_{?2}$ which book [$_{CP}$ do you think [$_{?1}$ which book [$_{CP}$ John bought which book]]]]

On the assumption that movement takes place cyclically, the wh-phrase *which book* moves as indicated by the arrows in (67). In this case, the movement cannot terminate at the edge of the embedded CP. This is predicted because the first internal merge creates $?_1$ = {NP, CP}, which fails to be labeled. A wh-phrase always ends up at the edge of a question sentence, and (67) is indeed fine when the wh-phrase moves on to the matrix-initial position. This is also expected. As the wh-phrase moves out of $?_1$, it can now be labeled by the CP it properly contains. The internal merge in the matrix clause creates $?_2$ = {NP, CP}. But in this case, NP is a wh-phrase and CP is a question. Then, they share the feature Q(uestion) and $?_2$ is labeled <Q, Q>.

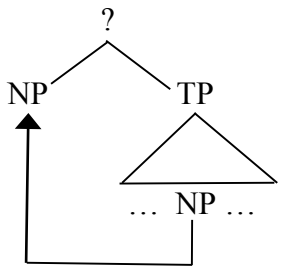
The hypothesis on labeling in Chomsky (2013) provides an independent explanation for the

last resort nature of internal Merge. Consider again (63b), repeated in (68).

(68) * Anything, the man did not eat _ .

As the negative polarity item *anything* does not qualify as a topic, the example is not an instance of topicalization. It is then simply internally-merged with the matrix TP as in (69).

(69)



But in this case, there is no feature sharing between the NP and TP, and consequently, ? cannot be labeled. Thus, (68) is excluded because of failure in labeling.

If this analysis is correct, last resort is not a property of internal Merge *per se*. Phrasal movement always creates an {XP, YP} structure. And there are two ways to label this. One is to move XP out of the structure. This is effective, but one cannot keep using this strategy forever as the movement must terminate at some point. And movement can terminate when it creates a configuration of feature sharing, the other case in which {XP, YP} can be labeled. This is the reason why movement appears to take place as the last resort.

4.2. Transformations and the Selectional Relations

There is another issue that arises when phrasal movement transformation is analyzed simply as *Merge*. Recall that it is argued in *Syntactic Structures* that selectional relations are represented at the pre-transformation structure. Phrase structure rules build the basic phrase structure that reflects selectional (thematic) relations, and transformational rules modify this basic structure. If external Merge and internal Merge apply in the parallel manner, a hypothesis can be entertained that NPs enter the structure at their selected positions by external Merge and then may move to non-thematic positions by internal Merge. In (66), for example, the subject NP is initially merged with *v*P at its thematic position and moves to the specifier position of TP, a non-thematic position. But the hypothesis implies that there is after all a distinction between external Merge and internal

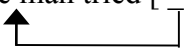
Merge as only the former can create configurations of selection.

However, arguments against this distinction have been presented in Hornstein (1989) and Bošković (2000), among others. They both argue that movement into thematic positions takes place. Hornstein, for example, first argues that there is good reason to allow this type of movement on conceptual grounds. In the LGB theory of Chomsky (1981), it is assumed just as in *Syntactic Structures* that selectional relations are represented at the pre-transformation structure, namely, D-structure. Hornstein points out that it followed from this assumption that a movement transformation cannot create a new selectional relation between a predicate and an argument. If it does, this new selectional relation was not represented at D-structure in contradiction with the assumption. However, once D-structure is abandoned, nothing blocks movement into a thematic position unless it is prohibited by stipulation. Then, it would be conceptually desirable to allow this type of movement.

Then, Hornstein goes on to reanalyze examples of control, such as (70a), in terms of movement as in (70b).

(70) a. The man tried [PRO to buy the book].

b. The man tried [_ to buy the book].



The man enters into selectional relation with the matrix verb *try* as the subject. Then, the LGB theory demands that it appears in the matrix subject position at D-structure. Since the embedded verb *buy* also selects for a subject, an independent element must appear in the position at D-structure. It was therefore assumed that a pronominal without phonetic content, PRO, appears in the position. However, Hornstein argues that given the elimination of D-structure, the example is best analyzed with movement as in (70b). *The man* is externally merged at the embedded subject position and receives the agent role there. Then, it internally merges at the matrix subject position, where it is interpreted as the agent of *try*.[9] If this analysis is on the right track, the distinction between external Merge and internal Merge is eliminated completely. Both freely apply and can create configurations that satisfy selectional requirements.

[9] More precisely, the NP is merged at the specifier position of *v*P and receives a theta role there in both the embedded and the matrix clauses. Independent support for this movement analysis of control can be found, for example, in Hornstein and Polinsky (2010).

The discussion above suggests that phrasal movement transformation, now called internal Merge, can be totally assimilated into *Merge*, a minimal operation that is required of Language. Then, it is clear why language has phrasal movement transformations. A language must have *Merge*, and phrasal movement comes with it.

5. Conclusion

Transformations were proposed in *Syntactic Structures* as they yield a simple, elegant theory. Further quest for simplicity and elegance led to the unification of phrasal movement transformations as a single principle *Move-α* on the relation between D-structure and S-structure. This, together with the elimination of D-structure and S-structure in the Minimalist program, made it possible to address deeper questions such as what transformations really are and why Language has them. Chomsky (1994) proposes that phrasal movement is an instance of the operation *Merge*, which forms a constituent of two elements. As discussed in the preceding section, efforts have been made to completely assimilate phrasal movements into *Merge*. If this project is successful, it identifies transformations, proposed on the basis of simplicity in *Syntactic Structures*, as instances of the minimal operation required of Language.

There was a great progress in the research on transformations in the last 60 years, and the Minimalist analysis entertained now is drastically simpler than its predecessors. And the progress undoubtedly will continue, as stated in the following passage from *Syntactic Structures*:

> Notice that neither the general theory nor the particular grammars are fixed for all time, in this view. Progress and revision may come from the discovery of new facts about particular languages, or from purely theoretical insights about organization of linguistic data – that is, new models for linguistic structure. … At any given time we can attempt to formulate as precisely as possible both the general theory and the set of associated grammars that must meet the empirical, external conditions of adequacy. (p.50)

That is, *Syntactic Structures* initiated research on syntax as a science, and the quest for simplicity will continue as long as syntax continues to be pursued as a science.

References

Baker, Mark, Kyle Johnson and Ian Roberts (1989) "Passive Arguments Raised," *Linguistic Inquiry* 20: 219-251.

Bošković, Željko (1997) *The Syntax of Nonfinite Complementation: An Economy Approach*, Cambridge, Mass.: MIT Press.

Chomsky, Noam (1957) *Syntactic Structures*, Mouton de Gruyter, Berlin.

Chomsky, Noam (1965) *Aspects of the Theory of Synt*ax, MIT Press, Cambridge, Mass.

Chomsky, Noam (1970) "Remarks on nominalization," in Roderick A. Jacobs and Peter S. Rosenbaum, eds., *Reading in English Transformational Grammar*, 184-221, Ginn and Co., Waltham, Mass.

Chomsky, Noam (1981) *Lectures on Government and Binding*, Foris Publications, Dordrecht.

Chomsky, Noam (1986a) *Barriers*, MIT Press, Cambridge, Mass.

Chomsky, Noam (1986b) *Knowledge of Language: Its Nature, Origin, and Use*, Praeger, New York.

Chomsky, Noam (1993) "A Minimalist Program for Linguistic Theory," in Kenneth Hale and Samuel Jay Keyser, eds., *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 1-52, MIT Press, Cambridge, Mass.

Chomsky, Noam (1994) "Bare Phrase Structure," in Gert Webelhuth, ed., *Government and Binding Theory and the Minimalist Program*, 383-439, Blackwell, Oxford.

Chomsky, Noam (1995) *The Minimalist Program*, MIT Press, Cambridge, Mass.

Chomsky, Noam (2013) "Problems of Projection," *Lingua* 130: 33-49.

Chomsky, Noam and Howard Lasnik (1977) "Filters and Control," *Linguistic Inquiry* 8: 425-504.

Collins, Chris (2005) "A Smuggling Approach to the Passive in English," *Syntax* 8: 81-120.

Hornstein, Norbert (1999) "Movement and Control," *Linguistic Inquiry* 30: 69-96.

Hornstein, Norbert and Maria Polinsky, eds. (2010) *Movement Theory of Control*, John Benjamins, Amsterdam.

Jaeggli, Osvaldo A. (1986) "Passive," *Linguistic Inquiry* 17: 587-622.

Koopman, Hilda and Dominique Sportiche (1991) "The Position of Subjects," *Lingua* 85: 211-258.

Lasnik, Howard (1981) "Restricting the Theory of Transformations: A Case Study," in Norbert Hornstein and David Lightfoot, eds., *Explanation in Linguistics: The Logical Problem of Language Acquisition*, 152-173, Longman, London.

Lasnik, Howard (1995) "Verbal Morphology: *Syntactic Structures* Meets the Minimalist Program," in Héctor Campos and Paula Kempchinsky, eds., *Evolution and Revolution in Linguistic Theory: Essays in Honor of Carlos Otero*, 251-275, Georgetown University Press, Washington, D.C.

Stowell, Timothy (1981) *Origins of Phrase Structure*, Ph.D. dissertation, MIT.