Creation myths of generative grammar, and the mathematics underlying *Syntactic Structures*[*]

Geoffrey K. Pullum
Linguistics and English Language
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Edinburgh EH8 9LL, Scotland, UK
gpullum@ling.ed.ac.uk

Noam Chomsky's extraordinarily influential book *Syntactic Structures* (Chomsky 1957; henceforth *SS*) inaugurated a fifty-year period during which the great majority of theoretical syntactic investigation done in the USA, and even worldwide, concentrated on a single underlying view of how syntax should be scientifically regarded. In the years after 1957, it was standard to take a sentence to be a sequence of words associated with some kind of description of its structure; to 'consider a *language* to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements'; and to take a grammar to be 'a device that generates all of the grammatical sequences ... and none of the ungrammatical ones' (*SS*: 13). Generating a sentence meant providing the possibility of constructing it by means of a procedure defined by the grammar, just as the arithmetical procedure 'take an integer and multiply it by itself' can be said to generate the set of perfect squares. Structural descriptions of sentences were taken to be derivational histories — sequences of steps through which a sentence could be derived.[1]

The impact of *SS* on American linguistics has often been compared to a scientific revolution in the sense of Kuhn. The proposals in the book convinced many linguists in the 1950s and early 1960s to turn their intellectual lives around and take up a new conception of their subject. Lasnik (2000) builds a graduate course in syntax on the content of *SS* together with some much more recent developments that are regarded as flowing directly from it. This paper takes a retrospective look at the descriptive, theoretical, and mathematical linguistic content of *SS*, and tries to assess the extent to which the proposals deserve the reverence that has been shown to them.

It is clear enough that certain creation myths about generative grammar have become widely prevalent in recent years. As a random example, consider the interview with Harvard psychologist and biologist Marc Hauser that was published in the May 2007 issue of *Discover* magazine on the occasion of the publication of his book *The Moral Mind*. In the course of introducing the conceptual background, interviewer Josie Glausiusz states that in order to develop the notion of 'a universal moral grammar within our brains' Hauser has drawn on 'the theories of MIT linguist Noam Chomsky, who in the 1950s proposed that all humans are equipped with a universal linguistic grammar, a set of instinctive rules that underlie all languages.'

But Chomsky never proposed any such thing in the 1950s, of course. It was not until the mid-1960s that he even spoke of 'universal grammar', and when he did he lost some of the people who had initially thought his work was a breakthrough. Yet it is extraordinarily common for science

[*]This paper was presented by invitation at the Mathematics of Language conference at UCLA in August 2007. This is the version of 12 June 2009. Some of the ideas presented here have their origin in collaborative work with Barbara Scholz, particularly the work that resulted in Scholz and Pullum (2007). I am very grateful to her for her generosity with assistance and advice — not that I have taken all of the advice. In addition I have benefited from discussions with Marcus Tomalin, who also presented a talk on *Syntactic Structures* at the UCLA meeting.

[1]To be more exact, equivalence classes of such sequences of steps in which the members differed only in the order of operations but not in the set of objects constructed.

journalists, and more generally for educated people who know a small amount of linguistics, to come away with the idea that Chomsky did indeed propose in the 1950s some kind of revolutionary theory of the mind and its 'instinctive rules', or that he discovered a rich array of universal laws that held for all languages.

Glausiusz and others have been encouraged in their beliefs by many interpreters and defenders of Chomsky's linguistics, for example David Lightfoot, whose introduction to the 2002 second edition of *SS* (not really a second edition, just a re-issue of the second printing of the first edition, with all the typographical errors of the original preserved in the almost entirely unrevised plates) begins thus:

(1) 'Noam Chomsky's *Syntactic Structures* was the snowball which began the avalanche of the modern "cognitive revolution". The cognitive perspective originated in the seventeenth century and now construes modern linguistics as part of psychology and human biology.' (Lightfoot 2002, v)

In truth there is not even a nod toward the study of cognition in *SS*, nor a mention or even an indirect flicker of interest in the Cartesian rationalism (or the Lockean empiricism) of the 17th century. Even the chapter on 'Goals of Linguistic Theory' has nothing about language acquisition or the nature of mind; it considers the relation of linguistic theory to the 'methods of analysis that an investigator might actually use, if he had the time, to construct a grammar from the raw data', and advocates the less ambitious idea that linguistic theory might provide 'a practical evaluation procedure for grammars' (p. 52), cautioning us against expecting it to inform us as to 'how, in princip[l]e, one might have arrived at the grammar of a language' (p. 56).

Tomalin (2006) is broadly correct in highlighting the role of nominalist philosophers such as Nelson Goodman, and Carnap's philosophy of science earlier than that, in the origins of *SS*. It would be a wild distortion to regard *SS* as a work on cognition, instinct, and the theoretical bases for developmental psycholinguistics. The revolutionary modernity of *SS* has been considerably overstated. This paper will for the most part not be concerned with such myth-making, but rather aims to examine the actual proposals made in the book.

1. Origin of generative grammars in the work of Emil Post The line of work that is of primary importance in understanding where the proposals of *SS* came from is, as Scholz and Pullum (2007) note, unfortunately missed by Tomalin (2006). Generative grammar springs from the mathematical and logical work of Emil Leon Post (1897–1954).

Beginning with his dissertation work in 1920, Post sought to reduce deduction in the propositional calculus to pure mathematics on strings — to eliminate any intuitive notion of logical consequence and replace it by purely formal operations on finite strings. He formalized rules of inference as operations that he called **productions**. His general definition of a production rule was given (at the beginning of Post 1943) in the following rather daunting tableau (I take the liberty of fixing a few small typographical errors in the original in *American Journal of Mathematics*, volume 65, 1943, p. 197):

(2)
$$
\begin{array}{l}
g_{1_1}\, P_{i'_1}\, g_{1_2}\, P_{i'_2}\, \cdots\, g_{1_{m_1}}\, P_{i'_{m_1}}\, g_{1_{m_1+1}} \\
g_{2_1}\, P_{i''_1}\, g_{2_2}\, P_{i''_2}\, \cdots\, g_{2_{m_2}}\, P_{i''_{m_2}}\, g_{2_{m_2+1}} \\
\ldots\ldots\ldots\ldots \\
g_{k_1}\, P_{i'^k_1}\, g_{k_2}\, P_{i'^k_2}\, \cdots\, g_{k_{m_k}}\, P_{i'^k_{m_k}}\, g_{k_{m_k+1}} \\
\text{produce} \\
g_1\, P_{i_1}\, g_2\, P_{i_2}\, \cdots\, g_m\, P_{i_m}\, g_{m+1}
\end{array}
$$

The g_i symbols in this array are metavariables; in particular instantiations of the schema are replaced by specific strings of symbols. The P_i are different: they are cover symbols, free variables over arbitrary sequences of symbols that might be present in the formulae to which the inference

rules are applied but which it is not specifically concerned with, except to leave the values unchanged. Thus an actual production might contain lines like vPw, or $P_1xP_2yP_3z$. We can call the set of strings before the word 'produces' the **left hand side** and the string after 'produces' the **right hand side**.

The index k in (2) identifies the number of premises in a particular instantiation of the schema, and the various m_i index the number of cover symbols in each line. Applying a rule of inference means matching a certain number of derived strings obtained so far in a proof to that number of lines in the left hand side, with substrings assigned to the cover symbols, and then computing the form of the new derived string by reference to the right hand side. The different strings do not all have to be of the same length (much of the forbidding typographical complexity stems from Post's careful attempt to indicate this, and to assign unique indices to every single symbol of every line).

Separately in his text, Post stipulates two restrictions on the form of productions. One is relatively unimportant, though necessary for technical reasons in a proof that he presents: although either the specific strings g_i or the cover symbols P_i in general may be assigned the null string as value, no identification of the cover symbols is permitted if it would lead to the right hand side of the production being instantiated as the empty string (i.e., it is required that the system should be unable to generate the null string as a conclusion). The other is more substantive and important. Just as no respectable rule of inference introduces arbitrary new material into a conclusion that was not in any of the premises, production systems do not allow arbitrary unspecified extra symbols to be added to the strings they derive. The cover symbols in the right hand side must all be repetitions of cover symbols found in the left hand side.

The idea, then, is that if under some assignment of values to the cover symbols all the strings represented by the members of the left hand side are legitimately derived from the axioms, then the string represented by the right hand side is also legitimately derived.

The rules of inference that one actually sees in logic texts usually (and mercifully) look quite simple compared to Post's schema. For example, take the rule known as Modus Ponens. It says that, for arbitrary subformulae φ and ψ, if we have derived a string of symbols looking like $((\varphi) \rightarrow (\psi))$ and we have also derived another that has the form (φ) (for the same value of φ), we are licensed to derive a string of the form (ψ). So that fits the schema in (2) if we let k (the number of premises) be 2 and let the second g_i be '$\rightarrow$'. We get this production as the formalization of Modus Ponens:

(3) $(P_1) \rightarrow (P_2)$
(P_1)
produces
(P_2)

This could be applied if we had derived both the string '$(p \wedge q) \rightarrow (r \vee \neg s)$' and the string '$(p \wedge q)$'. We would set $P_1 =$ '$p \wedge q$' and $P_2 =$ '$r \vee \neg s$', and the conclusion derived would be '$r \vee \neg s$'.

Post calls any system stated in the terms of his schema (2) a *canonical system*. A canonical system operating on a set of axioms generates the set consisting of all the axioms plus all the strings that can be obtained from the axioms by repeated application of the productions.

2. Post's work on generative capacity The major result of Post (1943) was that every set that can be generated by a canonical system can also be generated by a system in a vastly different format that Post calls *normal* form. A system is a *normal system* if it has just one axiom and all of its productions look like this:

(4) $x\, P$
produces
$P\, y$

More specifically, what Post proves is that given a canonical system Γ over an alphabet $A = \{a_1, a_2, \cdots, a_n\}$ and a second disjoint alphabet $B = \{B_1, B_2, \cdots, B_n\}$ (this extra alphabet being analogous to the linguist's nonterminals), we can construct a normal system Γ' over the alphabet $A \cup B$ such that the strings over A that are generated by Γ' are exactly the strings generated by Γ.

This brings out a very important point about Post's production systems from the standpoint of modern linguistics: The rewriting systems devised by Chomsky (1959) are just a special case of canonical systems with only a single one-symbol axiom.

Chomsky and Miller (1963:284) discuss from the ground up what grammatical rules could be like, and suggest that all rules of grammar will be of this form:

(5) $\phi_1, \ldots, \phi_n \rightarrow \phi_{n+1}$

'where each of the ϕ_i is a structure of some sort and where the relation $\rightarrow$ is to be interpreted as expressing the fact that if our process of recursive specification generates the structures $\phi_1, \ldots, \phi_n$ then it also generates the structure ϕ_{n+1}'. The correspondence to Post's more detailed (2) is obvious.

Although some of Chomsky's early works employed rules (generalized transformations) in which n was greater than 1, even the most liberal type of rule in the hierarchy he developed in Chomsky (1959), the one known as 'Type 0' rules, are just a highly restricted special case of Post's schema. Chomsky himself makes this observation:

(6) 'A rewriting rule is a special case of a production in the sense of Post; a rule of the form ZXW $\rightarrow$ ZYW, where Z or W (or both) may be null.' (Chomsky 1962:539)

To be specific about the special case here, the number of premises (the k index in (2)) is 1, the first and last specified strings in all lines of a production are null, and the number of cover symbols (the m index) is 2 for both left and right hand sides. In Post's notation, Type 0 rules look like this:

(7) $e\ P_1\ x\ P_2\ e$
produces
$e\ P_1\ y\ P_2\ e$

Chomsky uses a slightly different notation: Z and W are substrings used only as fixed context, X is used for the substring that is actually rewritten, Y is the new substring that the change introduces, and '$\rightarrow$' denotes the 'produces' relation; so in that notation the schema (7) would be written thus:

(8) $\ldots ZXW \ldots \rightarrow \ldots ZYW \ldots$

It would look even more familiar to phonologists if written in this form:

(9) $X \rightarrow Y / Z\_\_W$

Both (8) and (9) say that if a string of the form $\ldots ZXW \ldots$ can be legitimately obtained under the rule system, then the string $\ldots ZYW \ldots$ also counts as legitimately obtained.

Post had already considered rules of the type just considered. It was suggested to him by Alonzo Church that he tackle an open question that had been posed by Thue (1914): to find whether there was a decision procedure for determining whether a given string X could be converted into a given string Y by a set of rules of the form 'P_1WXZP_2 may be replaced by P_1WYZP_2, or conversely' (where W, X, Y, Z are strings over some fixed finite alphabet). Post (1947) answered the question by showing that even if Thue's bidirectional rules were replaced by unidirectional rules of the form $P_1WXYP_2 \rightarrow P_1WYZP_2$, which Post called 'semi-Thue' rules, there could be no such algorithm. He did this by proving that despite being a very limited special case of Post's canonical systems, Type 0 rule systems generate the full range of stringsets that canonical systems can generate. Just as in the case of Post's normal systems, the restriction to the Type 0 form does not make any otherwise generable set of strings impossible to generate — a set of semi-Thue rules

can enumerate any set that a canonical system can enumerate, and Post had already shown that there could be no decision algorithm for canonical systems.

The set of strings generated by either a canonical form system or a normal form system or a set of Type 0 rules is of course always a computably enumerable (c. e.) set, since the production system is a finite program for enumerating the membership; and every c. e. set is generable by some canonical system. By Post's theorem, it is also the case that every c. e. set is generable by some normal system.

So the bottom line is that not only do arbitary systems of rules of inference (formalized as Post productions) give us the power to describe (by generation) any set of strings that a Turing machine can recognize, but so do finite sets of rules saying simply 'delete X from the beginning of the current string and add Y on the end' for specified X and Y (Post's normal systems), and so do sets of rules that say 'replace X by Y when it occurs between W and Z' (Post's semi-Thue systems, and Chomsky's type 0).

Post's canonical form systems fall in with a whole slew of equivalent formalisms that may look very different but turn out to define exactly the same class of stringsets: Post's canonical form systems, normal form systems, Chomsky's Type 0 grammars, Turing machines (of many different modified forms), Church's lambda calculus, the μ-recursive functions of Herbrand and Gödel, and a variety of other formalisms all turn out to be equivalent ways of assigning finite descriptions to recursively enumerable sets.

Chomsky (1959: 137n) does credit the term 'generate' to Post, citing an informal paper on computably enumerable sets of positive integers that Post delivered as a lecture to the American Mathematical Society (Post, 1944), and he acknowledges Post (though without a bibliographical citation) in connection with the form of type 0 rewriting rules (1962: 539); but he appears never to have referred to any of Post's technical papers.[2] It is important that the idea of using abstract string-rewriting procedures as set-generating devices is due to Post, and Chomsky's early work continues on from that work seamlessly.

3. Infinitude of languages Chapter 2 of *SS* (page 13) begins with a definition of a language as a set of finite-length sentences over a finite vocabulary, and continues:

(10) 'All natural languages in their spoken or written form are languages in this sense, since each natural language has a finite number of phonemes (or letters in its alphabet) and each sentence is representable as a finite sequence of these phonemes (or letters), though there are infinitely many sentences.' (*SS*: 13)

The point that natural languages are infinite sets of finite strings became unquestionable dogma without ever being supported by serious argumentation. And almost five decades later, in *Language* (vol. 81, no. 1, 2005), we find Sam Epstein and Norbert Hornstein saying this in a letter:

(11) 'Compared to all other forms of animal communication ... human language is a highly structured formal combinatorial system and, in addition, the number of discrete well-formed sentences generated by the system is infinite.

This property of discrete infinity characterizes EVERY human language; none consist of a finite set of sentences. The unchanged central goal of linguistic theory over the last fifty years has been and remains to give a precise, formal characterization of this property and then to explain how humans develop (or grow) and use discretely infinite linguistic systems.'

Here infinitude is presented as an actual discovery, as if careful examination of all the attested human languages had revealed a denumerable infinity of sentences in each.[3] But the supposed infinite

[2] Alasdair Urquhart (personal communication) has suggested to me that this might be because his understanding of Post systems came from a secondary source, namely Rosenbloom (1950), which is cited in Chomsky (1956a) and Chomsky (1956b).

[3] There is a strange vacillation in this quotation concerning what 'system' means: in the first paragraph, what is

cardinality of languages is not a finding (see Scholz and Pullum 2009 on this point). Sentences of natural languages from the standpoint of linguists are somewhat like galaxies for astronomers: as we look for sentences of greater length or galaxies further away, we continue to find them. But while astronomers do not presume that this yields an answer to the question of whether the universe is infinite, linguists for some reason believe they can infer that the universe of sentences is infinite.

Chomsky deserves credit for seeing this clearly: *SS* does not proffer the fallacious idea that infinitude is an empirical finding. Chomsky states explicitly: 'In general, the assumption that languages are infinite is made in order to simplify the description of these languages' (pp. 23–24). This echoes the remarks of his mentor Zellig Harris: 'If we were to insist on a finite language, we would have to include in our grammar several highly arbitrary and numerical conditions' (Harris 1957: 208).

4. Probability and statistical approaches to grammaticality On page 17 of *SS*, after three short sections reviewing the possibility of glossing the predicate 'grammatical in English' as (i) belonging to the corpus of English, (ii) being meaningful to English speakers, or (iii) having a high probability of occurrence in an English-speaking community, Chomsky says:

(12) 'I think we are forced to conclude that grammar is autonomous and independent of meaning, and that probabilistic models give no particular insight into some of the basic problems of syntactic structure.' (*SS*: 17)

Here too, the stance of *SS* did not just become unquestionable doctrine in the form in which it was originally put; with respect to all three of the topics just raised it hardened into an ideological wall. For nearly fifty years generative grammarians have maintained rather extreme versions of Chomsky's positions on all three:

(13)
- Rightly rejecting the equation of grammaticality with corpus membership, they very largely fled to the opposite extreme and eschewed the use of corpus data completely, insisting that intuition should be the only approved source of evidence.
- Rightly stressing the importance of using purely syntactic criteria when they are appropriate, they wrongly gave syntactic accounts of what were in many cases quite clearly semantic phenomena, and rendered it controversial to propose any interweaving of semantic and syntactic constraints.
- Rightly distinguishing statistical evidence about the content of texts from syntactic evidence about structure, they went further, and shunned any reference to statistical evidence or probability theory for any purpose, failing to realize the potential inherent in stochastic models as they began to increase in importance within computational linguistics during the 1980s.

Chomsky discusses only very briefly the question of whether grammaticality might be reducible to probability of occurrence. In a certain sense, these remarks were a useful corrective to excessive enthusiasm for applications of Shannon/Weaver-style information theory to linguistics; but in another ways what he said was incorrect, and had damaging results on the development of the field.

Chomsky was right to dismiss Hockett's idea (Hockett, 1955) that the intuitive idea of ungrammaticality could be equated with the intuitive idea of being very unlikely to occur. There is a technical argument that he did not give; it was pointed out to me by Gerald Gazdar.

Take any grammatical possibility involving recursion, like (to take a simple case) adding *very* as a pre-head modifier to an adjective phrase (regardless of whether it already has an occurrence of

infinite is the number of sentences generated by the system, so the system is some kind of generative grammar, hence uncontroversially finite; but the second paragraph mentions 'infinite linguistic systems', suggesting that it is the system that is infinite. This may be just carelessness, or it may be due to an effort to echo of the strange conflation of languages with grammars that Chomsky has defended since about 1986 — another sign of Chomsky's extraordinary degree of influence.

very). In a probabilistic automaton there will be some probability associated with the transition, say $\frac{1}{k}$. Take any uncontroversially ungrammatical expression x — say, **the the cat*, and suppose it has a probability of occurrence $p(x) > 0$. (Ungrammatical expressions do occur in real texts; and in fact **the the cat* is not at all unlikely to occur, since people often stumble over articles and repeat them; but that is not essential, and we can take the probability of our chosen ungrammatical x to be arbitrarily low.) As the option of adding *very* is exercised, producing (by assumption) grammatical results, lower probabilities are obtained for the resultant phrase: whatever $p(\textit{good})$ might be, $p(\textit{very good}) = \frac{1}{k} \times n$, and $p(\textit{very very good}) = \frac{1}{k} \times \frac{1}{k} \times n$, and more generally, for any $q \geq 1$ a phrase of the form $\textit{very}^q$ *good* will have probability $\frac{1}{k^q} \times p(\textit{very good})$. But $\frac{1}{k^q}$ gets exponentially smaller as q gets larger, so for some value of q it has to be the case that $(\frac{1}{k^q} \times n) < p$. At that point that the putatively grammatical phrase $\textit{very}^q$ *good* will have lower probability than the ex-hypothesi ungrammatical phrase x.

In other words, in an infinite probabilistic language, probabilities for grammatical expressions can fall arbitrarily low, so no cutoff could ever be low enough to function as a surrogate for the grammaticality/ungrammaticality threshold and separate off the expressions that are definitely ill formed but have a non-zero likelihood of occurring. In that sense, Chomsky's intuition about probabilistic approximations to grammaticality was correct (though he gave no argument).

However, Chomsky also made a more specific claim, to the effect that probability could never distinguish a string like *Colorless green ideas sleep furiously* (fully grammatical though nonsensical) from a string like *Furiously sleep ideas green colorless* (entirely ungrammatical) turned out, when it was investigated four decades later, to be false. The strings two cannot be distinguished by any method like maximum likelihood estimation that which assigns zero as the probability of all events so far not observed. But methods incorporating a smoothing technique such as Good-Turing estimation can distinguish them, as Pereira (2000) showed. The probabilities of the two differ by at least five orders of magnitude according to Pereira's model, trained on newspaper text.

5. The proof that English is not finite-state It is very widely believed that a 'proof that English is finite-state' was given in *SS*. This is not true. Whether anything like the attempt made there can succeed is doubtful, but what is clear is that the 'proof' as Chomsky offered it in *SS* itself was just a few informal suggestions supporting the assertion that 'English is not a finite state language', that is, 'it is *impossible*, not just difficult, to construct a device of the [finite automaton] type ... which will produce all and only the grammatical sentences of English' (p. 23). Here the undergraduate level of the exposition is particularly clear. Grammars are not distinguished from accepting automata; finite-state Markov processes are not distinguished from their transition graphs ('state diagrams'); no definitions are given.

Indeed, it is not entirely clear that Chomsky had a good intuitive grasp of the richness and complexity of the class of finite-state languages. The example he gives of an infinite language that is finite-state is this one:

(14) **the old***((**man comes**) + (**men come**))

But this is not just finite-state; it is also star-free, and k-locally testable for all $k \geq 2$, and strictly k-local for all $k \geq 2$. Here is a strict SL_2 grammar for it (I use '⋊' and '⋉' for the beginning-of-string and end-of-string markers):

(15) $\{\langle ⋊ \textbf{ the}\rangle, \langle \textbf{the old}\rangle, \langle \textbf{old old}\rangle, \langle \textbf{old man}\rangle, \langle \textbf{old men}\rangle, \langle \textbf{man comes}\rangle, \langle \textbf{men come}\rangle,$
$\langle \textbf{comes } ⋉\rangle, \langle \textbf{come } ⋉\rangle\}$

Notice that the quasi-recursive device used in Chomsky's example is just a self-loop in the transition diagram that permits iteration on a single element; there is no use of return to repeat internally complex structural sequences.

It is interesting that there is this indication of failure to distinguish strictly local from finite-state (or anything else in the subregular hierarchies), because it emerges again in Bever et al. (1968),

where associationist psychology is linked to a language-theoretic notion that appears implicitly to be strict SL_2 but is not expositorily distinguished from finite-state.

Chomsky's full argument that natural languages are not finite-state was, of course, offered in a celebrated technical paper of the year before: Chomsky (1956b), cited in *SS* on p. 22. This contains the 'rigorous proof' to which *SS* alludes on p. 23. But if that argument is sound, to my knowledge no one has established the fact. Certainly I can report that I do not understand it.

The argument depends on ternary relation of 'm-dependency' that holds between a string S, an integer m, and a language L over a vocabulary A. Chomsky defines a string S as having an m-dependency with respect to a stringset $L \subseteq A^*$ iff S has this form:

(16) $S = x_1 a_1 x_2 a_2 \ldots x_m a_m z x_1 b_1 x_1 b_2 x_m b_m$

and there is a unique permutation of the numbers $(1, \ldots, m)$ — a mapping α from that set to itself — meeting this condition (which I quote from Chomsky 1956b):

(17) 'there are $\{c_1, \ldots, c_{2m} \in A$ such that for each subsequence $(i_1, \ldots, i_p)$ of $(1, \ldots m)$, S_1 is not a sentence of L and S_2 is a sentence of L, where

(10) S_1 is formed by substituting c_{i_j} for a_{i_j} in S, for each $j \leq p$; S_2 is formed by substituting $c_{m+\alpha(i)}$ for $b_{\alpha(i_j)}$ in S_1, for each $j \leq p$.'

So the changed string S_1 has had some changes made to the a symbols, with the result that S_1 is not in L, and only by changing the b symbols at the corresponding positions, in a way determined by the mapping α, can membership in L be restored.

Chomsky now asserts: 'Evidently, if S has an m-dependency with respect to L, at least 2^m states are necessary in the finite-state grammar that generates the language L.' No proof of this lower bound is offered, and Lars Svenonius (1957) states that the claim is not true: he says m states will suffice (and will still support Chomsky's proof), though he does not offer any argument or discussion. I do not know which claim is correct. (It seems not to matter much. Chomsky needs the number that he claims is 2^m merely so that he can claim that for any finite-state language there must be a finite upper bound k such that no sentence has an n-dependency for any $n > k$.)

What is important is not the mathematical unclarities here, but the point established by Daly (1974), namely that the formal language theory claim Chomsky states cannot be easily related to his natural language material. That material involves pairs like ⟨*if*, *then*⟩, ⟨*both*, *and*⟩, ⟨*either*, *or*⟩, ⟨*neither*, *nor*⟩. Notice that it just is not true that if you take one of these pairs in a text and replace its second member by the second member of one of the other pairs you can only restore grammaticality by replacing the analogous substitution on the first member. It might seem that way from a first impression, but it is not. Consider (18):

(18) a. *If$_i$ neither$_j$ he nor$_j$ I can do it then$_i$ someone else will.*

b. **Neither if he nor I can do it then someone else will.*

c. **Neither if he then I can do it nor someone else will.*

It will become ungrammatical if we switch *if* and *neither* (we get (18b)), and we cannot fix things by switching *nor* and *then* (we get (18c)). Perhaps this is not the intended application of the statement in (17); but I cannot find a more appropriate one. Daly (1974) spends many pages attempting to work out how there could be a sound argument for Chomsky's conclusion based on m-dependencies, and he does not find one: he leaves it unresolved whether the argument can be constructed.

There is an additional mathematical shortcoming in what Chomsky says in *SS*. He remarks in summing up: 'Thus we can find various kinds of non-finite state models within English' (*SS*: 22–23). But this does not suggest any appreciation of the fact that finite-state stringsets can contain infinite non-finite-state subsets. His presentation reads as if discovering some non-finite-state proper subset of English would suffice to show that English was not finite-state. This is not true, of course,

unless the proper subset in question can be extracted by some regularity-preserving operation like homomorphism or intersection with a regular set. With regard to establishing this for English, it is important that all of the words in the pairs above (*if*, *then*, *both*, *and*, *either*, *or*, *neither*, *nor*) can occur in sentences without the other member of the pair. It is not clear that there is *any* item β in English such that if $\varphi\beta\psi$ is in English then $\psi = \psi_1\gamma\psi_2$ for $\psi_1 \neq e$ and $\gamma \neq \beta$.

6. The attack on context-free grammars If natural languages (whether regular or not) were all context-free (CF), the development of transformations would have vastly less motivation than if there were non-CF natural languages. Today it is assumed that there are natural languages with non-CF stringsets: the Swiss German of the Zurich area seems to be one (Shieber, 1985). But Chomsky did not even hint in *SS* at any argument that natural languages might be beyond the weak generative capacity limits of context-free phrase structure grammars (CF-PSGs).

He does argue, however, that CF-PSGs are inadequate. His arguments are purely intuitive ones, based on personal intuitions about of elegance in description. And they depend on three main topics: coordination, the auxiliary system, and passive clauses. I now turn to an examination of those.

7. The coordination principle The generalization that I will call the coordination principle is the one stated in *SS* as formula (26) on p. 36:

(19) Chomsky's coordination principle
'If S_1 and S_2 are grammatical sentences, and S_1 differs from S_2 only in that X appears in S_1 where Y appears in S_2 (i.e., $S_1 = \ldots X \ldots$ and $S_2 = \ldots Y \ldots$), and X and Y are constituents of the same type in S_1 and S_2, respectively, then S_3 is a sentence, where S_3 is the result of replacing X by $X +$ *and* $+ Y$ in S_1 (i.e., $S_3 = \ldots X +$ *and* $+ Y \ldots$).'

I note that S_1 and S_2 are required to be 'grammatical sentences', i.e., they are generated by the grammar — a grammar of which (19) itself is a part. In later work Chomsky insisted that it was a naive view to think of transformations applying to sentences in the language; they applied to abstract strings of formatives that underly sentences. But he certainly says 'sentences' in the quotation above. In fact the way he phrases the coordination principle involves existential quantification over the entire content of the language, in the manner of what were later to be called transderivational constraints.

It seems to me that this way of formulating things raises the possibility of paradox, or at least failure of groundedness. If we were to choose S_1 and S_2 so that they are grammatical only if (19) is true, would (19) be true?

Presumably the idea is that we are designing a recursive definition, in the way that Bar-Hillel (1953) called 'recursive in disguise'. It would be something like this:

(20) Preliminary recursive definition of 'grammatical sentence':

— If an output of the phrase structure rules undergoes (in the proper sequence) all the singulary transformations that are marked obligatory and any subset of the singulary transformations that are marked optional, then the result is a grammatical sentence.

— If there are S_1 and S_2 which are grammatical sentences and S_3 is formed as detailed in (19), then S_3 is a grammatical sentence.

$\cdots$ [OTHER SUCH CONDITIONS HERE]

— Nothing else is a grammatical sentence.

But reapplying the second step to sentences whose formation has already employed that step demands knowing the categories that figure in those sentences, and that is not provided. A considerable increase in explicitness is called for.

Later the coordination is restated in a way that is supposed to be more explicit, as a 'generalized transformation'; but it is a remarkable fact that it seems less explicit, not more. Here is the reformulation:

(21) Conjunction transformation from *SS*:

Structural analysis: of S_1: $Z - X - W$
of S_2: $Z - X - W$

where X is a minimal element (e.g., *NP*, *VP*, etc.) and Z, W are segments of terminal strings.

Structural change: $(X_1 - X_2 - X_3; X_4 - X_5 - X_6) \rightarrow$
$X_1 - X_2 + \textit{and} + X_5 - X_3$

The '*S*' in the variable names 'S_1' and 'S_2' might suggest 'sentence', but S_1 and S_2 are not sentences; they are what formal language theorists call sentential forms, strings representing possible stages in phrase structure derivations.

X is stipulated to be a 'minimal element', but this term is undefined. From the clarificatory example it appears to mean 'single symbol in the non-terminal vocabulary'. And in fact X seems to be a variable over non-terminals that is constrained to have the same value in S_1 as it has in S_2 (because if X could be chosen differently in S_1 and S_2, coordinating of arbitrary distinct categories would be allowed).

Z and W are stipulated to be 'segments of terminal strings.' I take this to mean that they are substrings of terminal-symbol strings that the grammar generates. So if the generated language is $L \subseteq V_T^*$, there are strings $\psi_1, \ldots, \psi_4 \in V_T^*$ such that $\psi_1 Z \psi_2 \in L$ and $\psi_3 W \psi_4 \in L$. And it is clear that Z and W are variables that must take the same values in S_1 and S_2 (otherwise they would be redundant: S_1 and S_2 could be any arbitrary strings containing an instance of X).

This means the analyses of S_1 and S_2 are completely identical: Z, X, and W all have to be the same. An example would be something like $S_1 = S_2 =$ *Put NP in the truck*. And now we see that nowhere in (21) is it stated that the terminal strings of the two instances of the category X have to be distinct. As it is given, we could choose:

(22) $X =$ *it*: *Put it in the truck.*
$Y =$ *it*: *Put it in the truck.*
$X +$ *and* $+ Y =$ *it and it*:
**Put it and it in the truck.*

This is presumably unintentional: (19) mentions that 'S_1 differs from S_2', but (21) does not, so there is more content in the informal coordination principle than in its formalized version.

Let us assume, then, that Chomsky intended to say that S_1 and S_2 are supposed to be identical sentential forms of distinct generated terminal strings. That is fully statable in the formalisms of *LSLT*, though it is fairly complicated.

Now notice that no real point turns on making reference to both of S_1 and S_2. The structural analysis part of the rule is saying merely that we are concerned with a grammatical simple clause S that can be broken down a string of words Z at the beginning, a string of words W at the end, and in the middle a string of words that belongs to some category X (which therefore is allowed to follow Z and precede W). No further use is made of Z and W in the statement of the rule.

In fact the 'structural change' of the rule not only throws away Z and W, it throws away X as well. None of these variables are mentioned again. Instead, six new variables $X_1, \ldots, X_6$ are introduced. The X in these variable names has no relation to the prior use of X. We are not told what they range over, and no mapping is given that connects them to Z and X and W. If careful readers of *SS* were not puzzled by this, they were not paying attention.

We are left to guess that $X_1, \ldots, X_6$ range over terminal strings, and that $X_1 = X_4 = Z$, and $X_3 = X_6 = W$, and $X_2 \neq X_4$, and both X_2 and X_4 bear the 'is a' relation to X (i.e., there is a legitimate derivation in which $\varphi_1 X \psi_1 \overset{+}{\underset{G}{\Rightarrow}} \varphi_1 X_2 \psi_1$ and another in which $\varphi_2 X \psi_2 \overset{+}{\underset{G}{\Rightarrow}} \varphi_2 X_4 \psi_2$, for $\varphi_1, \varphi_2, \psi_1, \psi_2 \in (V_N \cup V_T)^*$). None of this is made explicit in (21). Nine variables have been used to hold four values (the terminal strings Z, X_2, and W, and the category X), and they have not been explicitly related.

This is a turgid and inexpert deployment of mathematical symbolism. In the end the content of the rule appears to be specifiable much more simply. A nonterminal symbol X can be replaced by

the string 'X *and* X' in any context whatsoever. A simple phrase structure rule '$X \rightarrow X$ *and* X' could do that. And such a rule would say something else, something on which the *SS* account is silent: it would assign a phrase of the form 'X *and* X' to a specific category, namely X. From *SS* we get no hint concerning whether coordinations are phrases of the same syntactic type as their coordinate constituents, or phrases of some distinct syntactic type (which would make it unclear how coordinates of an NP can themselves be coordinations), or perhaps phrases of a maximally unspecified type that pick up their syntactic properties from their immediate constituents (as, implicitly, in the account offered by Huddleston et al. 2005, Ch 14). Strings like *I went to the supermarket and to the bank this morning* are (under a generous interpretation of the under-explained formalism) generated, but no answer is given to the question of whether *to the supermarket and to the bank* is a constituent of PP type, or a constituent at all.

Conceivably the idea of using a rule *NP* $\rightarrow$ *NPNP* to generate NP-coordinations was ruled out because it does not permit the reconstruction of unique trees from phrase structure derivations (see McCawley 1968), but this is conjecture as far as I can see.

Note that nothing is said in *SS* about multiple coordination. We know we can generate sentences like *It was indigo and violet*, but we cannot generate *The conventional rainbow colors are red, orange, yellow, green, blue, indigo, and violet*. Can a single generalized transformation reapply iteratively to its own output? How is it that some of the coordinators disappear? Whence comes the constraint saying that the coordinator (*and* or some other coordinator word) can be placed only before the last coordinate? And so on.

Added to all this is a real empirical problem with the *SS* account: the generalization that the coordination principle (19) and the rule (21) are intended to express — we will rely on the informal statement in (19) rather than the bungled formalization in (21) — is wildly false. I will supply a number of examples.

(23) PREDICATIVE COMPLEMENTS AND DIRECT OBJECTS

X = *became*: *He <u>became</u> a fine teacher.*
Y = *admired*: *He <u>admired</u> a fine teacher.*
X + *and* + Y = <u>*became*</u> *and* <u>*admired*</u>:
**He <u>became and admired</u> a fine teacher.*

(24) ADJUNCTS AND COMPLEMENTS

X = *last year*: *He left <u>last year</u> without a word.*
Y = *his three children*: *He left <u>his three children</u> without a word.*
X + *and* + Y = <u>*last year*</u> *and* <u>*his three children*</u>:
**He left <u>last year and his three children</u> without a word.*

(25) EXPANDED COORDINATES

X = *or Pat*: *Either Kim <u>or Pat</u> should do it.*
Y = *or Bob*: *Either Kim <u>or Bob</u> should do it.*
X + *and* + Y = <u>*or Pat*</u> *and* <u>*or Bob*</u>:
**Either Kim <u>or Pat and or Bob</u> should do it.*

(26) PHRASES INTRODUCED BY *for*

X = *for nobody else would*: *We had to act, <u>for nobody else would</u>.*
Y = *for time was short*: *We had to act, <u>for time was short</u>.*
X + *and* + Y = *for nobody else would and for time was short*:
**We had to act, <u>for nobody else would and for time was short</u>.*

(27) PHRASES INTRODUCED BY RESULTATIVE *so*

X = *so I went to bed*: *I was tired <u>so I went to bed</u>.*

Y = *so I turned up my iPod*: *I was tired so I turned up my iPod.*
X + *and* + Y = *so I went to bed and so I turned up my iPod*:
**I was tired, so I went to bed and so I turned up my iPod.*

(28) PHRASES INTRODUCED BY CONNECTIVE *only*

X = *only I was out of town*: *I would have helped, only I was out of town.*
Y = *only you never called*: *I would have helped, only they never called.*
X + *and* + Y = *only I was out of town and only they never called*:
**I would have helped, only I was out of town and only they never called.*

(29) DETERMINERS

X = *my*: *Have you seen my paper?*
Y = *this*: *Have you seen this paper?*
X + *and* + Y = *my and this*:
**Have you seen my and this paper?*

(30) VERB AGREEMENT CONTROLLERS

X = *Don*: *I think Don was the soloist.*
Y = *Phil*: *I think Phil was the soloist.*
X + *and* + Y = *Don and Phil*:
**I think Don and Phil was the soloist.*

The bottom line here is that *SS* does not provide a careful description of coordination. Its formalization is unsuccessful — there is actually more clarity in the informal (19) than there is in the supposed formal version (21) — and empirically the description achieves very little and in empirical terms is falsified by large collections of familiar facts.

Certainly, 1957 was very early in the modern history of syntactic investigation, and we look back with the advantage of fifty years of hindsight; but it is worth reminding ourselves that there is only a very sketchy and imperfect contribution to understanding coordination here.

8. The analysis of auxiliaries The *SS* analysis of the English auxiliaries is often taken to represent an original breakthrough. But in fact its first leading idea — as embodied in the famous phrase structure rule

(31) $Aux \rightarrow C(M)(have + en)(be + ing)(be + en)$

— largely coincides with the analysis given by Charles C. Fries, *The Structure of English* (1952). Fries's proposal is presented like this:

(32)

GROUP	CLASS	GROUP				CLASS
A	**1**	**B**				**2**
		(a)	(b)	(c)	(d)	
The	students	*may*	*have*	*had to*	*be*	moving
T	N	Aux				V

Notice that elements like *may*, perfect *have*, progressive *be*, etc., are not being treated as verbs (Class 2 words) here. They are nonverb dependents, grouped into multiple columns under the general heading of 'GROUP **B**', a subsidiary category covering the function words associated with verbs.

This is a variant of what Huddleston and Pullum (2002) calls the **dependent-auxiliary** analysis. It treats the auxiliary verbs as non-verbal morphemes that cluster together as a non-endocentric pseudo-constituent.

Chomsky's phrase structure rules for the auxiliary system are stated as follows:

Sentence $\rightarrow$ *NP* + *VP*
VP $\rightarrow$ *Verb* + *NP*
Verb $\rightarrow$ *Aux* + *V*
Aux $\rightarrow C(M)$(*have* + *en*)

Elements like *may*, *will*, perfect *have*, progressive *be*, etc., could be regarded 'verbs' in some informal sense, but they are not treated as verbs in this analysis — not in any sense.

These rules mesh crucially with a transformation often known as 'Affix Hopping', but referred to in *SS* as the Auxiliary Transformation, and formulated thus:

(33) Auxiliary Transformation — obligatory:

Structural analysis: *X* – *Af* – *v* – *Y* (where *Af* is any *C* or is *en* or *ing*;
v is any *M* or *V*, or *have* or *be*)

Structural change: $X_1 - X_2 - X_3 - X_4 \rightarrow$
$X_1 - X_3 - X_2 \# - X_4$

The 'Affix Hopping' component of the *SS* analysis won many converts to TG. But it was beset with very significant problems — mostly not recognized at the time.

It is initially quite hard to understand. For example, it uses a bewildering array of possibilities for what might correspond to the informal notion of 'verb'; indeed, it is is replete with terminological and notational equivocations and inconsistencies in this connection, using no less than six different competing symbols in the grammar, all with mnemonically motivated names that appear to be designed to suggest 'verb'.

- *Verb* is introduced as a lexical node on p. 28, but then is clearly treated as a phrasal node on p. 39.
- V is introduced as a lexical node on p. 39, and is equated with 'verb' on p. 42, but then becomes a phrasal node on p.79 (where *consider a fool* is analyzed as a V).
- v is an informal cover symbol embracing two elements that would be traditionally interpreted as either verb lexemes or verb stems (*have* and *be*) together with the category M of modals and the category V.
- V_1 is apparently the lexical category of a class of verbs like bring on pp. 76-77.
- V_a is apparently the lexical category of a class of verbs like consider on pp. 76-77.
- V_2 is apparently the lexical category of a class of verbs like consider on p. 112 (this may be just a notational inconsistency introduced in the process of writing the book).

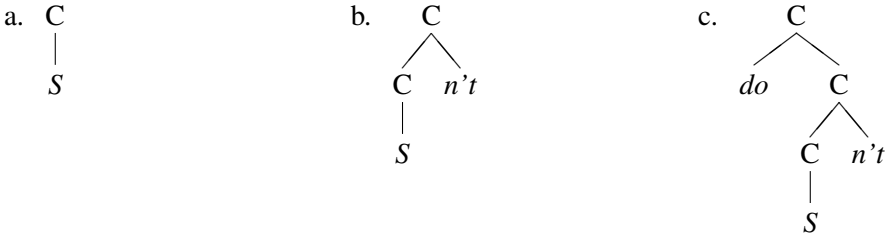
I note also that *Aux* is referred to as the 'auxiliary phrase', hence apparently a phrasal node, on p. 42; but it is called the 'auxiliary verb', suggesting a lexical category, on p. 43. It actually appears that *SS* is attempting to analyze the whole of the syntax of English auxiliary verbs without making any reference to the notion 'auxiliary verb', for nothing in the *SS* analysis corresponds to that notion. The *Aux* node is really just an artificial housing for a cluster of elements that are represented as being non-verb non-heads sisters; and *Verb* is a node for grouping *Aux* together with one lexical verb (the unit that some descriptions referred to as the 'verbal group'). And it is an odd constituent: an non-endocentric phrase containing half a dozen daughter elements but no head, and a phrasal node to which no transformation ever applies: *Aux* is never moved, deleted, inserted, or targeted by adjunction.

It is entirely unclear to me how the *SS* analysis of auxiliaries came to be regarded as elegant or attractive. I never regarded it thus. It seems to me to have a host of quite serious disadvantages. Some emerge as disadvantages given later widely accepted advances in syntactic analysis, but many are not anachronistic in this way, and should have been apparent at the time.

Affix Hopping is illegal One of the most serious objections in theoretical terms is that the analysis is simply not compatible with formal theory of *LSLT*: as noted by Sampson (1979), the Auxiliary Transformation is not a legal transformation at all under the theory of *LSLT*.

Syntactic structure of *doesn't* Another objection is that this is a syntactic transformation that affects internal structure of words, which syntax generally doesn't do. Moreover, in doing so it sometimes builds structures that are linguistically quite implausible. As Zwicky (1969) first pointed out, the structure assigned to the first word of *Doesn't it fit?* is singularly strange. When no auxiliaries are present, the optional negation transformation T_{not} attaches the negative formative *n't* to a C element, which has been realized by the morpheme S according to the preceding transformation, the Number Transformation, so we get the structure diagrammed in (34a).

(34)



Then T_q (subject-auxiliary inversion) can apply to form (34b), and finally the formative *do* attached to yield (34c). Because *n't* has to be attached to the tense/number/person element C before supportive ***do*** is inserted, *does* turns out not to be a constituent in *doesn't*; instead, *sn't* is a constituent!

Ordering paradoxes Next, recall that Chomsky in 1957 took transformations to apply in a fixed linear order determined by stipulation within the grammar. At least four fairly serious ordering paradoxes emerge in the *SS* analysis.

Ordering paradox 1: The Auxiliary Transformation must apply before VP ellipsis in John has (so that the ·*en* suffix disappears along with the elided VP), but it must apply after VP ellipsis in *John did* (so that the tense/number/person element C gets is stranded). There is no order that gets the right results.

Ordering paradox 2: The T*not* transformation that attaches the negative element to auxiliaries must apply before Subject-Auxiliary Inversion (SAI) so that we can derive *Haven't they?*; but SAI must precede the Do-Transformation so we can get *Did they?*; and the Do-Transformation must be able to precede T*not* so we can get *They don't.* Again there is no consistent order in which these rules can be arranged in the grammar.

Ordering paradox 3: As just noted, T*not* must apply before SAI to get *Haven't they?*, and SAI must precede the Auxiliary Transformation to get *Did they?* (rather than *Do they went?). Hence by transitivity of ordering, T*not* precedes the Auxiliary Transformation. But that means *They hadn't* must go through a stage where it is *PAST – have + n't*, where *haven't* is a unit; and that means that the Auxiliary Transformation will produce *They haven'ted.

Ordering paradox 4: Quantifier Floating (QF) must precede SAI to get *Have the others all been swimming?* (rather than *Have the others been all swimming?*), and SAI precedes the Auxiliary Transformation, so by transitivity QF precedes the Auxiliary Transformation; but the Auxiliary Transformation must precede QF, otherwise QF will block it: *all the others PAST go swimming* ⇒ *the others PAST all go swimming* ⇒ **The others did all go swimming.* (Read this without emphatic stress on *do*.)

These may not be considered problematic today; but they were major defects given the assumptions of the 1950s.

9. Equivocation about 'verb' Here is another problem, not hitherto noted as far as I am aware. The *SS* analysis of the verb-particle construction uses a rule $V \rightarrow V_1\ Prt$ (see p.75). But which

constituent is now the verb for purposes of the Auxiliary Transformation's reference to 'any M or V, or *have* or *be*'?

If V is the verb for the Auxiliary Transformation, the suffix will follow the particle, and we will get **The police bring inned the criminal.*

If V_1 is the relevant category and is syntactically distinct from V, the structural analysis for the Auxiliary Transformation will not be met (v covers M or V or *have* or *be*, but not V_1).

And if V_1 is the relevant category and the difference in notation is purely expository, so that syntactically $V_1 = V$, then the Auxiliary Transformation ignores a fundamental universal of syntax that Chomsky developed in subsequent years, the A-over-A principle. In any of these three cases, there is a real problem with verb-particle constructions.

Inversion of *have* On p. 66 of *SS* it is pointed out that *John hasn't a chance* and *Has John a chance?* can be derived (alongside John doesn't have a chance and Does John have a chance?) because T_{not} is optional; but that also gets arbitrarily many bad strings:

(35) a. They had that old Morris chair recovered.
b. ∗They hadn't that old Morris chair recovered.
c. ∗Had they that old Morris chair recovered?

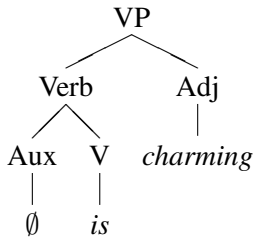
(36) a. Rex had Priscilla three times on their wedding night.
b. ∗Rex hadn't Priscilla three times on their wedding night.
c. ∗Had Rex Priscilla three times on their wedding night?

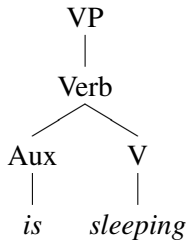
(37) a. Your dog has rabies.
b. ∗Your dog hasn't rabies.
c. ∗Has your dog rabies?

(38) a. The manager had the staff set things up.
b. ∗The manager hadn't the staff set things up.
c. ∗Had the manager the staff set things up?

(39) a. The couple had sex several times.
b. ∗The couple hadn't sex several times.
c. ∗Had the couple sex several times?

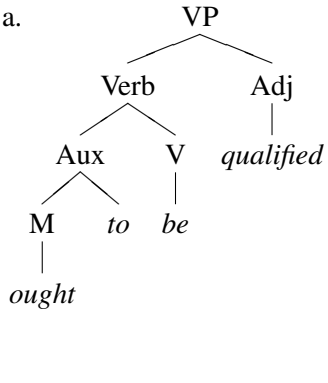
Branching structure of catenative sequences It is an intuitively puzzling feature of the grammar proposed in *SS* that it assigns such different phrase structures would be assigned to *is charming* and *is sleeping*:

(40) a.


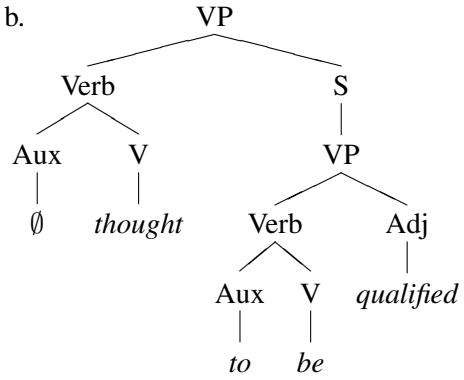
b.


And *ought to be overseas* will have an entirely different phrase structure from *thought to be overseas*. This is because ***ought*** belongs in the category M (it does not take the 3rd singular present tense suffix *·s*, it does take *n't*, it appears only in tensed contexts, it precedes perfect *have* and progressive *be*, and it does invert), while ***think*** is a lexical verb — a V, therefore — taking a clausal complement. So the two structures are:

(41) a.



b.

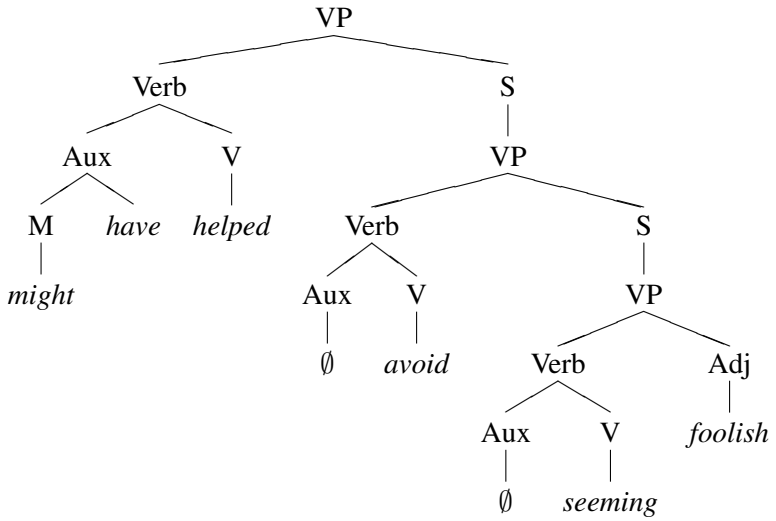


Notice that a sequence like *might have helped avoid seeming aggressive* could in principle be bracketed in a very large number of ways, and Chomsky has (apparently arbitrarily) settled for a very strange one:

(42) [[[*might have*] *helped*] [*avoid* [*seeming* [*foolish*]]]]

The tree representation of the surface structure that would apparently result from the *SS* grammar is this:

(43)



Here *helped avoid* is analyzed as a verb taking a subjectless (and auxiliary-free) complement clause, and is produced in an entirely different way (via a generalized transformation of embedding) from *have helped*, which is not a verb-complement construction at all. This is an arbitrary distinction without syntactic motivation. As is well known, VP ellipsis phenomena and many other considerations argue for a uniformly right-branching structure (such as is defended by Huddleston and Pullum 2002, of course).

The unneeded Aux node If assumed in modern analyses, *Aux* would often block desired c-command relations (Pullum and Wilson 1977 cite several cases of this). It would guarantee, for example, that no tense or agreement element ever c-commands the V of a clause.

And of course modern analyses do not employ the *Aux* node or anything analogous. All of the items formerly housed in *Aux* are now treated as heads of projections.

This is just as was always recommended by proponents of the primary alternative to the dependent-auxiliary analysis. That alternative has been presented in many minor variants over the years, going back to classic accounts like that of Jespersen (who referred to the modals as the 'anomalous finites' in the verb system), and defended by such works as Ross (1967), McCawley (1971), McCawley (1975), Newmeyer (1975), Pullum and Wilson (1977), Gazdar et al. (1982), Huddleston (1974), Huddleston (1976), and many others.

The specific version that *CGEL* adopts is termed the **catenative-auxiliary analysis** (see *CGEL*, Ch 14, §4.2, pp. 1209ff). What it has in common with the accounts like those of Pullum and Wilson (1977) or Gazdar et al. (1982) is that it analyzes the auxiliaries of English as verbs (belonging to a syntactic subclass with certain special behaviors) that take complements in the same way that other complements do. More specifically, *CGEL* analyzes auxiliaries as verbs that take **catenative** complements: non-finite, VP-internal, subjectless complements that are neither direct objects nor predicative complements, capable of recursive embedding leading to chains of verbs (*hoped to seem to tend to want to avoid appearing to have been thought to expect to...*, etc.).

The specific order of auxiliaries (modal, perfect *have*, progressive *be*) do not need to be specified by a phrase structure rule. Briefly (for this is old stuff, explained much more carefully in such works as Ross 1967, McCawley 1971, McCawley 1975, Newmeyer 1975, Pullum and Wilson 1977, and Gazdar et al. (1982)):

- any modal has to come first because modals only have primary forms and thus can only head tensed VPs;
- there can be no second modal because modals select bare infinitival complements, and bare infinitival complements have plain-form head verbs, and modals have no plain forms;
- perfect *have* cannot follow progressive *be* because the former is a grammaticized item expressing secondary tense and its semantic contribution is strongly incompatible with progressive aspectuality.[4]

The issue between the rival analyses of auxiliaries becomes important in the next section.

10. The analysis of passive clauses The analysis of passive clauses in *SS* is motivated by reference to four unpleasant problems that allegedly emerge if passives are treated with phrase structure rules.

1. *Verb* expands as *Aux* – *V*, and the element *be* + *en* can be selected under *Aux* only if the *V* is transitive, and stating this would complicate the rule system.

2. Even if *V* is transitive, *be* + *en* cannot be selected if *V* is followed by *NP* — another condition to state.

3. If *V* is followed by the PP *by* + *NP*, then *be* + *en* is obligatory in *Aux* — a third complex co-occurrence that has to be built into the rules.

4. Selection restrictions reverse: active subjects ∼ *by*-phrase objects; passive subjects ∼ direct objects.

But all four of these claims are spurious.

Claim 1: It is not true that *be* + *en* occurs only with transitives:

(44) *Man is descended from apes.* (compare **Somebody descended man from apes*).

(45) *Charles is rumored to be dissatisfied* (compare **Somebody rumored Charles to be dissatisfied*).

(46) *Lord Howe Island was uninhabited until recent times* (compare **People uninhabited Lord Howe Island until recent times*).

Claim 2: It is not true that *be* + *en* is forbidden if there is a following NP:

[4] In rough outline, *is walking* denotes the property of being temporally located in the midst of a ongoing walking episode that is extended in time. Given that *has walked* denotes being temporally located at a present point with the completion of a completed walk lying at an earlier point, the phrase *is having walked* would have an utterly incoherent meaning.

(47) *I've often been called an idiot.*

(48) *He was denied all his legal rights.*

(49) *We were shown several nice apartments.*

Claim 3: It is not true that if a *by*-phrase is present *be* + *en* is required. The many refuting examples include all the clauses that *CGEL* calls **bare passives** (the bracketed part of *We had this* [*done by an expert*], for example, or *He went and got himself* [*stung by a wasp*]) and **concealed passives** (like *This car wants* [*cleaning*] or *The book needs* [*revising by an experienced editor*]).

Claim 4: It is not true that syntax has to tackle the selection restriction issues, as has been clear since McCawley's classic 1968 paper 'Concerning the base component of a transformational grammar'. Chomsky's work sustained such a level of suspicion about semantics that the question of whether the grammar of English should distinguish *John plays golf* from *Golf plays John* (the latter being referred to as a 'non-sentence') was answered unhesitatingly in the affirmative. But this cannot be right. As McCawley convinced us, consideration of a range of examples suggests that every semantic property a noun phrase could have would be capable of playing a role in a selection restriction.

I would argue, in fact, that something stronger is true: that selection restrictions were never not linguistic at all. Consider the frame in (50a), and consider the results of filling the blank with each the nouns in (50b):

(50) a. *This _____ thinks someone is going to give it some food.*

b. *child*, *toddler*, *baby*, *bonobo*, *tamarin*, *dog*, *kitten*, *rat*, *echidna*, *chicken*, *crocodile*, *rattlesnake*, *frog*, *shark*, *halibut*, *lobster*, *cockroach*, *oyster*, *jellyfish*, *nematode*, *lily*, *bacterium*, *virus*, *rock* ...

Certainly, beyond some point as we go along the list of nouns from beginning to end we will reach some that render the constructed sentence bizarre, in virtue of the incongruity of the subject and the verb ***think***; but different people will have different cutoff points at which the bizarreness starts. One can imagine significant debates about the right cutoff point taking place among, say, philosophers of mind or animal rights advocates. But what one cannot imagine is that the correct way to settle the matter is through descriptive English syntax. Whatever factors determine the right answer (if there is a right answer), they are surely not linguistic factors.

Or to put it another way (suggested to me by Barbara Scholz), there has been controversy in philosophy at least since Turing's famous 1950 paper in mind about the issue of whether machines might some day be able to think; but surely that philosophical issue is not to be decided by research in linguistics — the identification of the maximally simple correct generative grammar for English.

All four of the arguments given for the passive transformation, then, are entirely unpersuasive. The fourth just is a conceptual mistake, and the other three are artifacts of a bad analysis of the auxiliary system.

It is important that the right analysis of the auxiliary system provides for the right analysis of passives too. Auxiliary verbs take **catenative complements**: non-finite subjectless recursively nestable clauses with specified inflectional features. Different verbs are lexically specified for different kinds of complement:

- ***think***: declarative content clause with an optional subordinator *that*;
- ***inquire***: interrogative content clause with obligatory subordinator (*whether* or *if*) or preposed *wh*-phrase;
- ***tend*** (and the modal ***ought***): subjectless *to*-infinitival catenative complement clause;
- most modal verbs, and auxiliary ***do***: bare infinitival catenative complement clause;
- progressive ***be***: gerund-participial clause;

– perfect ***have***: past-participial clause;
– ***be***, and ***get***: bare passive catenative complement clause, with past-participial inflection on the head verb;
– causative ***have***: direct object NP plus bare passive catenative complement clause; ...

The key feature of passive clauses that gives them their special usefulness is that their meanings employ the sense of the verb in a way that involves what might be called role reversal: instead of the VP denoting a property of the agent, it denotes a property of the patient. But the latter — the semantics where it is a property of the patient that is denoted by the VP — is not at all tied to the elements present in the *SS* Passive Transformation.

– it is not tied to the presence of *be* + *en*, as bare passives show;
– it is not tied to the presence of *be* + *en*, as concealed passives show;
– it is not tied to the presence of an immediately postverbal NP, as prepositional passives show;
– it is not tied to the existence of a corresponding active clause, as passives with verbs like *rumored* and *said* show;
– and it is not tied to clauses, as we see from the ambiguity of *the shooting of the hunters*.

One clear fact about the *SS* analysis of passive clauses is that the generalization expressed in the advocated transformational rule is plainly and massively false. The rule entails very clearly that for any *NP* immediately after any V, a grammatical passive will be produced if that postverbal *NP* is shifted to subject position and the original subject is shifted into a *by*-phrase and *be* is added before the head verb and the head verb is inflected in past-participial form. There are indefinitely many counterexamples, of many interestingly different types. Bach (1980) gives a significant number. Postal (2004) has catalogued a much large number — though even his list is not exhaustive.

Postal's interest lies in pursuing a rather exaggeratedly personal allegation of dishonesty on Chomsky's part for repeatedly saying that passivization applies blindly to all $V-NP$ sequences that , and that is not my concern here. Rather, I think (in agreement with Bach) that the sheer richness of the array of unpassivizable $NP-Aux-V-NP$ sequences is of interest because it tells us something significant about the passive construction: it is very highly lexical, and the notion that it represents some kind of simple, automatic, regular, syntactic modification process has no plausibility whatsoever. Here, in random order, are a representative fifty examples of $NP-Aux-V-NP$ sequences that do not yield well-formed results under the *SS* passive transformation. (In order to focus on what seem to be genuine lexical or semantic exceptions to passivization, and to allow for the fact that *SS* assumed that complement clauses were transformationally embedded rather than base-generated as complements, I do not bother to give any examples of the indefinitely vast class in which the postverbal NP is the subject of an internal content-clause complement: it is true that *The bishop hopes you will agree* yields **You are hoped will agree by the bishop* under the *SS* passive transformation, but I ignore cases of that sort.)

(51) Fifty active clauses that do not passivize

Mike seemed a nice enough guy.	**A nice enough guy was seemed by Mike.*
I felt such a fool.	**Such a fool was felt by me.*
I doubt if it mattered a lot to him.	**I doubt if a lot was mattered to him by it.*
You looked a bit of a fop in that tie.	**A bit of a fop was looked in that tie by you.*
The suspect fled the scene.	**The scene was fled by the suspect.*
That cost me a lot of money.	**I was cost a lot of money by that.*
The boss had the staff write it up.	**The staff were had write it up by the boss.*
He's done me so many favors.	**I've been done so many favors by him.*
CGELweighs nearly six pounds.	**Nearly six pounds are weighed by CGEL.*
The train couldn't approach the station.	**The station couldn't be approached by the train.*
An appetizer began the five-course dinner.	**The five-course dinner was begun by an appetizer.*
The remark betrayed contempt for science.	**Contempt for science was betrayed by the remark.*

The dean didn't buy my argument.	**My argument wasn't bought by the dean.*
Armand caught the flu.	**The flu was caught by Armand.*
The man croaked something unintelligible.	**Something unintelligible was croaked by the man.*
The train departed the station at 11 A.M.	**The station was departed at 11 A.M. by the train.*
The abbott muttered something bitter.	**Something bitter was muttered by the abbott.*
Fred lacked all finesse.	**All finesse was lacked by Fred.*
The key wouldn't enter the lock.	**The lock wouldn't be entered by the key.*
My theory fits the facts.	**The facts are fit by my theory.*
Snakes don't give milk.	**Milk is not given by snakes.*
Horace heard that from Mildred.	**That was heard from Mildred by Horace.*
The cabinet includes the Home Secretary.	**The Home Secretary is included by the cabinet.*
Such behavior inspires loathing.	**Loathing is inspired by such behavior.*
The affair involved foreign banks.	**Foreign banks were involved by the affair.*
US 95 will lead you to New Haven.	**You will be led to New Haven by US 95.*
Ellen left Chicago in June.	**Chicago was left by Ellen in June.*
Evelyn doesn't mind profanity.	**Profanity isn't minded by Evelyn.*
The Titanic neared the iceberg.	**The iceberg was neared by the Titanic.*
The decision permitted me to remain in class.	**I was permitted to remain in class by the decision.*
Hugh quit the police.	**The police was quit by Hugh.*
The package never reached Gwen.	**Gwen was never reached by the package.*
Your dog has rabies.	**Rabies is had by your dog.*
Carmen can't stand sushi.	**Sushi can't be stood by Carmen.*
Some can't afford adequate health care.	**Adequate health care can't be afforded by some.*
The kids couldn't tell the depth.	**The depth couldn't be told by the kids.*
Tom wanted pizza.	**Pizza was wanted by Tom.*
The experiment yielded a strange result.	**A strange result was yielded by the experiment.*
Mary never answered Greg.	**Greg was never answered by Mary.*
The general never cabled Louisa.	**Louisa was never cabled by the general.*
Her name eludes me.	**I am eluded by her name.*
My dear uncle departed this life last May.	**This life was departed last May by my dear uncle.*
The legal system had failed her.	**She had been failed by the legal system.*
2007 found the US still at war.	**The US was found still at war by 2007.*
The truth just hit me.	**I was just hit by the truth.*
A six-pack doesn't last Mark long.	**Mark isn't lasted long by a six-pack.*
That movie starred Madonna.	**Madonna was starred by that movie.*
Laura struck everyone as smart.	**Everyone was struck as smart by Laura.*
That did not suit the dean.	**The dean was not suited by that.*
Jim never wrote Irene.	**Irene was never written by Jim.*

But it is also the case that even if we set aside the huge number of cases in which the *SS* generalization fails we find that a large number of instances of what are intuitively passives are not covered by the transformation given. I do not think it has been adequately appreciated just how many distinct passive constructions there are in English. (This may be one of the reasons why those word processors with a built-in grammar checker that is supposed to flag all passive clauses, according to the usual writing-teacher prejudice against them, are so bad at it.) By my count, we are dealing with two dozen distinct constructions.

Here is a brief overview. Passive clauses such as *liked by his classmates* or *beaten down by her troubles* or *irritated by his kids* are non-finite clauses that have distributions not very different from adjective phrases such as *popular with his classmates* or *weary from her troubles* or *angry with his kids*. They can be found as complements of ascriptive uses of the copula (compare *was liked by his classmates* and *was popular with his classmates*), or in various simple intransitive constructions (compare *looked beaten down by her troubles* and *looked weary from her troubles*), or in various complex-transitive constructions (compare *got irritated by his kids* and *got angry with his kids*).

(52) a. *He was well liked by his classmates.* [passive VP]

b. *He was decidedly popular with his classmates.* [AdjP]

(53) a. *She looked beaten down by her troubles.* [passive VP]

b. *She looked weary from her troubles.* [AdjP]

(54) a. *I often got irritated by his kids.* [passive VP]

b. *I often got angry with his kids.* [AdjP]

They may be inflected in past-participal form (as in the ones in the above examples) or in some cases gerund-participal form (the 'concealed passive', as in *The book merits re-reading*). Some of them are adjectival passives with obligatory *un-*. And cross-cutting these distinctions almost completely are the lines dividing prepositional passives (often called pseudo-passives, with stranded prepositions) from the ordinary kind, and separating long passives (with the *by*-phrase complement) from short passive clauses (without it). The full array of constructions we get looks like this:

(55) The 24 passive constructions of English

a. long non-prepositional ascriptive-copular passive (LNA):
Lucy was examined by the doctor.

b. short non-prepositional ascriptive-copular passive (SNA):
Lucy was examined.

c. long prepositional ascriptive-copular passive (LPA):
Lucy was looked at by the doctor.

d. short prepositional ascriptive-copular passive (SPA):
Lucy was looked at.

e. long non-prepositional simple-intransitive passive (LNI):
Lucy got examined by the doctor.

f. short non-prepositional simple-intransitive passive (SNI):
Lucy got examined.

g. long prepositional simple-intransitive passive (LPI):
Lucy got looked at by the doctor.

h. short prepositional simple-intransitive passive (SPI):
Lucy got looked at.

i. long non-prepositional complex-transitive passive (LNT):
The parents had Lucy examined by the doctor.

j. short non-prepositional complex-transitive passive (SNT):
The parents had Lucy examined.

k. long prepositional complex-transitive passive (LPT):
The parents had Lucy looked at by the doctor.

l. short prepositional complex-transitive passive (SPT):
The parents had Lucy looked at.

m. long non-prepositional *be un*-adjectival passive (LNBU):
Lucy was unexamined by the doctor.

n. short non-prepositional *be un*-adjectival passive (SNBU):
Lucy was unexamined.

o. long prepositional *be un*-adjectival passive (LPBU):
Lucy was unlooked at by the doctor.

p. short prepositional *be un*-adjectival passive (SPBU):
Lucy was unlooked at.

q. long non-prepositional *go un*-adjectival passive (LNGU):
Lucy went unexamined by the doctor.

r. short non-prepositional *go un*-adjectival passive (SNGU):
Lucy went unexamined.

s. long prepositional *go un*-adjectival passive (LPGU):
Lucy went unlooked at by the doctor.

t. short prepositional *go un*-adjectival passive (SPGU):
Lucy went unlooked at.

u. long non-prepositional concealed passive (LNC):
Lucy needs examining by the doctor.

v. short non-prepositional concealed passive (SNC):
Lucy needs examining.

w. long prepositional concealed passive (LPC):
Lucy needs looking at by the doctor.

x. short prepositional concealed passive (SPC):
Lucy needs looking at.

Notice that the *SS* transformation handles just one (the first) out of these 24. But it has no special priority or importance relative to the others. If it expressed a true generalization (which it does not), it would be expressing a generalization holding over only a very small part of the range inherent in the descriptive task of characterizing English passive clauses.

I conclude, then, that *SS* exhibited very little motivation for transformations. Its proposed claims and analyses had grave shortcomings.

11. Conclusions

SS had enormous influence on theoretical linguistics. It ushered in fifty years of generative grammatical research, and today it is rare indeed to find a department in the United States that does not include at least some professors who do generative linguistic research. Why was this, when the arguments it gave were unsound at so many points, and so many of its claims were directly refutable and subsequently refuted?

Quite probably the answer has to do with Chomsky's stirring call for a greater abstractness in linguistic theorizing and a broader scientific view of what linguistics was about. It should never be forgotten that this call was heard, and a new subfield sprang into existence in response — it came to be known as theoretical linguistics, which may be something of a disservice to the highly theoretical bent of earlier investigators like Zellig Harris, but is understandable nonetheless. It is that reorientation of the field that is the legacy of *SS*, and it atones for what I would argue is the incompetent and slapdash symbolization and the grossly inaccurate description to be found in the book's main content chapters.

Somehow Chomsky persuaded a large part of the field to follow the classic advice that adults often find themselves forced to supply to children: 'Do as I say, not as I do.' He did not in fact apply the lessons of recursive function theory to grammar, or pay proper attention to the solvability (decidability) question, or demonstrate rigorous formalization of grammatical rules, or provide careful proofs of his claims, or establish language-theoretic claims about natural languages, or show that natural languages are infinite, or propose a theory of universal grammar or of language acquisition, or revolutionize the study of the mind. He just convinced large numbers of other people that they should attempt these things. For that he certainly deserves our gratitude, and *SS* deserves our respect; but those who represent *SS* as actually having achieved rather than proposed all the foregoing ambitions (Lasnik 2000, to take a particularly extreme case) are perpetuating an error about the history of our field.

References

BACH, EMMON. 1980. In defense of passive. Linguistics and Philosophy 3. 297–341.

BAR-HILLEL, YEHOSHUA. 1953. On recursive definitions in empirical science. Actes du XIème Congrès Internationale de Philosophie, Volume V: Logique, analyse philosophique, philosophie des mathématiques, 160–165. Amsterdam: North-Holland.

BEVER, THOMAS G.; JERRY A. FODOR; and MERRILL GARRETT. 1968. A formal limitation of associationism. Verbal Behavior and General Behavior Theory, edited by Theodore R. Dixon and David L. Horton, 582–585. Englewood Cliffs, NJ: Prentice-Hall.

CHOMSKY, NOAM. 1956a. The Logical Structure of Linguistic Theory. Cambridge, MA: MIT Library. Microfilmed; revised version of a 1955 unpublished manuscript.

CHOMSKY, NOAM. 1956b. Three models for the description of language. I.R.E. Transactions on Information Theory IT-2. 113–123. Reprinted in *Readings in Mathematical Psychology*, Volume II, ed. by R. Duncan Luce, Robert R. Bush, and Eugene Galanter, 105–124; New York: John Wiley & Sons, 1965.

CHOMSKY, NOAM. 1957. Syntactic Structures. The Hague: Mouton.

CHOMSKY, NOAM. 1959. On certain formal properties of grammars. Information and Control 2. 137–167. Reprinted in *Readings in Mathematical Psychology*, Volume II, ed. by R. Duncan Luce, Robert R. Bush, and Eugene Galanter, 125–155, New York: John Wiley & Sons, 1965 (citation to the original on p. 125 of this reprinting is incorrect).

CHOMSKY, NOAM. 1962. Explanatory models in linguistics. Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress, edited by Ernest Nagel; Patrick Suppes; and Alfred Tarski, 528–550. Stanford, CA: Stanford University Press.

CHOMSKY, NOAM; and GEORGE A. MILLER. 1963. Introduction to the formal analysis of natural languages. Handbook of Mathematical Psychology, edited by R. Duncan Luce; Robert R. Bush; and Eugene Galanter, vol. II, 269–321. New York: John Wiley and Sons.

DALY, ROBERT TIMON. 1974. Applications of the Mathematical Theory of Linguistics. The Hague: Mouton.

GAZDAR, GERALD; GEOFFREY K. PULLUM; and IVAN A. SAG. 1982. Auxiliaries and related phenomena in a restrictive theory of grammar. Language 58. 591–638.

HARRIS, ZELLIG. 1957. Co-occurrence and transformation in linguistic structure. Language 33. 283–340. Published version of a paper read to the Linguistic Society of America in 1955. Page reference is to the reprinting in Jerry A. Fodor and Jerrold J. Katz (eds.), *The Structure of Language: Readings in the Philosophy of Language*, 155–210 (Englewood Cliffs: Prentice-Hall).

HOCKETT, CHARLES F. 1955. A Manual of Phonology. No. 21: 4, Part 1, Memoir 11 in International Journal of American Linguistics. Bloomington, IN: Indiana University.

HUDDLESTON, RODNEY. 1974. Further remarks on the analysis of auxiliaries as main verbs. Foundations of Language 11. 215–229.

HUDDLESTON, RODNEY. 1976. An Introduction to English Transformational Syntax. London: Longman.

HUDDLESTON, RODNEY; and GEOFFREY K. PULLUM. 2002. The Cambridge Grammar of the English Language. Cambridge: Cambridge University Press.

HUDDLESTON, RODNEY; GEOFFREY K. PULLUM; ET AL. 2005. A Student's Grammar of the English Language. Cambridge: Cambridge University Press.

LASNIK, HOWARD. 2000. *Syntactic Structures* Revisited: Contemporary Lectures on Classic Transformational Theory. Cambridge, MA: MIT Press.

LIGHTFOOT, DAVID. 2002. Introduction. In Noam Chomsky, *Syntactic Structures*, second edition, v–xviii. Mouton de Gruyter, Berlin.

MCCAWLEY, JAMES D. 1968. Concerning the base component of a transformational grammar. Foundations of Language 4. 243–269. Reprinted in James D. McCawley, *Grammar and Meaning*, 35–58 (New York: Academic Press; Tokyo: Taishukan, 1973).

MCCAWLEY, JAMES D. 1971. Tense and time reference in english. Studies in Linguistic Semantics, edited by Charles J. Fillmore and D. Terence Langendoen, 97–113. New York: Holt, Rinehart and Winston.

MCCAWLEY, JAMES D. 1975. The category status of English modals. Foundations of Language 12. 597–601.

NEWMEYER, FREDERICK J. 1975. English Aspectual Verbs. The Hague: Mouton.

PEREIRA, FERNANDO. 2000. Formal grammar and information theory: Together again? Philosophical Transactions of the Royal Society 358. 1239–1253.

POST, EMIL. 1943. Formal reductions of the general combinatory decision problem. American Journal of Mathematics 65. 197–215.

POST, EMIL. 1944. Recursively enumerable sets of positive integers and their decision problems. Bulletin of the American Mathematical Society 50. 284–316.

POST, EMIL. 1947. Recursive unsolvability of a problem of thue. Journal of Symbolic Logic 12. 1–11.

POSTAL, PAUL M. 2004. Skeptical Linguistic Essays. New York: Oxford University Press.

PULLUM, GEOFFREY; and DEIRDRE WILSON. 1977. Autonomous syntax and the analysis of auxiliaries. Language 53. 741–788.

ROSENBLOOM, PAUL. 1950. The Elements of Mathematical Logic. New York: Dover.

ROSS, JOHN ROBERT. 1967. Auxiliaries as main verbs. Studies in Philosophical Linguistics 1. 77–102.

SAMPSON, GEOFFREY. 1979. What was transformational grammar? Lingua 48. 355–378.

SCHOLZ, BARBARA C.; and GEOFFREY K. PULLUM. 2007. Tracking the origins of transformational generative grammar. Journal of Linguistics 43. 701–723.

SCHOLZ, BARBARA C.; and GEOFFREY K. PULLUM. 2009. Recursion and the infinitude claim. Recursion in Human Language, edited by Harry van der Hulst, ??? Berlin: Mouton de Gruyter.

SHIEBER, STUART. 1985. Evidence against the context-freeness of human language. Linguistics and Philosophy 8. 333–343.

SVENONIUS, LARS. 1957. Review of 'three models for the description of language' by noam chomsky. Journal of Symbolic Logic 23. 71–72.

THUE, AXEL. 1914. Probleme über veränderungen von zeichenreihen nach gegebenen regeln. Skrifter utgit av Videnskapsselskapet i Kristiana, I, no. 10 in Matematisk-naturvidenskabelig klasse 1914. Oslo: Norske Videnskaps-Akademi.

TOMALIN, MARCUS. 2006. Linguistics and the Formal Sciences. Cambridge: Cambridge University Press.