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

Geoffrey K. Pullum[1]

School of Philosophy, Psychology and Language Sciences, University of Edinburgh
3 Charles Street, Edinburgh EH8 9AD, UK
gpullum@ling.ed.ac.uk
WWW home page: http://ling.ed.ac.uk/~gpullum/

Abstract. *Syntactic Structures* (Chomsky [6]) is widely believed to have laid the foundations of a cognitive revolution in linguistic science, and to have presented (i) the first use in linguistics of powerful new ideas regarding grammars as generative systems, (ii) a proof that English was not a regular language, (iii) decisive syntactic arguments against context-free phrase structure grammar description, and (iv) a demonstration of how transformational rules could provide a formal solution to those problems. None of these things are true. This paper offers a retrospective analysis and evaluation.

1 Introduction

Syntactic Structures (Chomsky [6], henceforth *SS*) was not just another contribution to the discipline of structural linguistics. In the opinion of many American linguists, it ended the structuralist period. Martin Joos's definitive anthology of structuralist work *Readings in Linguistics I* first appeared in the same year, and it now looks more like an obituary than a reader. The study of syntax was altered forever by the introduction in *SS* of transformational generative grammar (TGG). Forty years later, Howard Lasnik's introductory graduate syntax course at the University of Connecticut was still built around the content of *SS* together with more recent developments that he regarded as flowing directly from it (see Lasnik [20]).

But people have come to believe things about *SS* that were never true. Some linguists encourage such false beliefs. Lightfoot [21] opens his introduction to the 'second edition' of *SS* (actually just a re-issue of the second printing of the first edition, retaining the typographical errors) by stating that 'Noam Chomsky's *Syntactic Structures* was the snowball which began the avalanche of the modern "cognitive revolution"... [which] originated in the seventeenth century and now

⋆ This paper is based on an invited presentation at the Mathematics of Language conference at UCLA in August 2007. Many of the ideas here have been profitably discussed with my collaborator Barbara Scholz. I am very grateful to her for her generosity with assistance and advice — not that I have taken all of the advice.

construes modern linguistics as part of psychology and human biology.' There was not even a nod toward the study of cognition in *SS*, nor a flicker of interest in the 17th century. Lightfoot's psychobiological snowball is just an invention.

In this paper I try to counter some of the myth-making about *SS*, focusing on the mathematical bases for the statement of grammars rather than any anachronistic claims about the philosophical origins or cognitive implications of the proposals in *SS*. I begin by examining the origins of the conception of grammars that *SS* introduced.

2 Generative grammar and the work of Emil Post

TGG originates in work that was aimed at mathematicizing logical proof. Above all it stems from early work by the Polish-American mathematical logician Emil Leon Post (1897–1954).

2.1 Production systems

SS defines 'the form of grammar associated with the theory of linguistic structure based upon constituent analysis' thus (*SS*, p. 29):

> Each such grammar is defined by a finite set Σ of initial strings and a finite set F of 'instruction formulas' of the form $X \to Y$ interpreted: "rewrite X as Y." Though X need not be a single symbol, only a single symbol of X can be rewritten in forming Y.

As an example, Chomsky gives a grammar where $\Sigma = \{Z\}$ and F contains the rules $Z \to ab$ and $Z \to aZb$. The stringset generated is $\{a^n b^n | n \geq 1\}$. Chomsky adds (p. 31):

> It is important to observe that in describing this language we have introduced a symbol Z which is not contained in the sentences of this language. This is the essential fact about phrase structure which gives it its 'abstract' character.

It will be clear to anyone acquainted with Emil Post's mathematical work that a grammar of the sort Chomsky has defined is a special case of what Post called a production system.

Post started out trying to formalize the logic informally assumed in Whitehead and Russell in *Principia Mathematica*, and ended up with a characterization of the recursively enumerable (r. e.) sets. He formalized inference rules as **productions**. A production associates a set of given strings (the **premises**) to a new string (the **conclusion**), which the premises are said to 'produce'.

A **production system** consists of a set of initial strings (this corresponds to the Σ of *SS*) and a set of productions (corresponding to the set F of 'instruction formulas' in *SS*). (Post [28] is the definitive journal article; Brainerd & Landweber [3] provides a very useful elementary exposition with worked examples.) Given a set $\{\phi_1, \ldots, \phi_n\}$ of initial strings and/or strings derived from

them by the productions (where $n \geq 1$), a production saying '$\{\phi_1, \ldots, \phi_n\}$ produces ϕ_{n+1}' legitimates the addition of ϕ_{n+1} to the collection of strings that are derived or generated.

Twenty years after Post [28], Chomsky and Miller [12] propose (p. 284) that rules of grammar are of this form:

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

They explain: 'each of the ϕ_i is a structure of some sort and ... 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}.' Clearly, they might just as well have said that they take grammatical rules to be productions in the sense of Post [28].

Generative capacity

However, Post did more than simply invent what were later to be called generative grammars. He also proved the first theorems concerning generative capacity. The major result of Post [28] was a theorem concerning the expressive power of production systems with a radically limited format for productions.

Post's original definition of productions was maximally general, with no limits on number or complexity of premises. The ϕ_i are of the form $g_0 P_1 g_1 P_2 \ldots g_{k-1} P_k g_k$ (for $k \geq 0$), where the g_i are specified constant strings of symbols over a vocabulary Ω and the P_i are free variables that can take any string in Ω^* as value, and carry it over to the conclusion if that variable appears there. Post called these maximally general production systems 'canonical systems', but he proved that the same generative power was obtainable with productions of a much simpler form.

Normal systems The main theorem of Post [28] is that every set generated by a canonical system can also be generated by a system in a much more restricted format called a 'normal system'. In a normal system there is just one axiom, and all productions take the form '$g_1\ P$ produces $P\ g_2$', where P is a free variable and g_1 and g_2 are specified strings. To be more precise, Post's theorem is this:

(2) **Theorem** (Post [28]) Given a canonical system Γ over a finite vocabulary Ω_T it is always possible to construct a normal system Γ' over $\Omega = \Omega_T \cup \Omega_N$ (where Ω_N is a new set of symbols disjoint from Ω_T) such that Γ' generates $x \in \Omega_T^*$ iff Γ generates x.

This shows that a radical limitation on rule form, restricting rules to saying 'Any string beginning with g_1 may be rewritten with its g_1 prefix erased and g_2 added at the end', has no effect at all on generative capacity.

The extra symbols in Ω_N that do not appear in generated strings are of course the ones that Chomsky described as essential to the abstract character of phrase structure: they are the symbols he would later call nonterminals.

Semi-Thue systems There is another specially limited form of productions. Chomsky [9] calls these 'rewriting rules', and recognizes explicitly that they are restricted forms of Post's production systems:

> A rewriting rule is a special case of a production in the sense of Post; a rule of the form ZXW → ZYW, where Z or W (or both) may be null. (Chomsky [9]: 539)

Productions in this format were called **type-0 rules** in Chomsky [7]. The number of premises is limited to 1, and all of W, X, Y, Z are specified strings. The only free variables are the flanking ones covering whatever precedes W and whatever follows Z. Thus in Post's notation such as rule would say '$P_1g_1g_2g_3P_2$ produces $P_1g_1g_4g_3P_2$'. This replaces g_2 by g_4 if g_1 immediately precedes and g_3 immediately follows.

This restriction originates in a technical paper from ten years before in which Post (following a suggestion by Alonzo Church) tackled an open question posed by Axel Thue [40]. Thue had asked whether there was a decision procedure for determining whether a specified string X could be converted into a given string Y by a set of rules of the form '$WXZ \leftrightarrow WYZ$, where W, X, Y, Z are strings over some fixed finite alphabet and $\phi \leftrightarrow \psi$ is to be read as 'ϕ may be replaced by ψ or conversely'.

Post [30] answers Thue's question by showing first that if there is a decision procedure for Thue-style bidirectional systems (where for every $\phi \rightarrow \psi$ we also have the inverse $\psi \rightarrow \phi$) there is a decision procedure for unidirectional ones (which do not necessarily have the inverses), and this is known not to be true, so the reduction shows that the decision problem for Thue systems—the type-0 rules of Chomsky—is recursively unsolvable.

2.2 Recursive enumerability

Post had thus proved the first two theorems in what would later come to be known as the theory of generative power of grammars. Both of his results show that radical limitations on the form of rules may have no effect on what can be generated. The importance of Chomsky [7] was that it showed other restrictions did limit what could be generated (for example, '$P_1g_1g_2g_3P_2$ produces $P_1g_1g_4g_3P_2$' with the restriction that $|g_2| \leq |g_4|$ will generate only context-sensitive stringsets). But the transformations introduced in *SS* did not entail any such limitations. Hilary Putnam, in a remarkably prescient paper [33], discussed his reasons for thinking that natural languages had to have a decidable membership problem, and then remarked:

> Chomsky's general characterization of a transformational grammar is much too wide. It is easy to show that any recursively enumerable set of sentences could be generated by a transformational grammar in Chomsky's sense.

He provided no proof, but his conclusion was surely correct. There were no signs of limitations on the form of transformations that could restrict their expressive power more tightly than that of canonical systems.

There was one element that Chomsky added to production systems in developing generative grammars: the device of 'extrinsic' rule ordering. He required that a grammar should define a strict ordering on its rules, so that each rule R_i would be permitted to apply (if at all) only after all the rules ordered before it had applied, and before any of the rules ordered after it had applied. But this had no restrictive effect on generative power. No one ever offered an example of a stringset that can be generated by some unordered set of productions but cannot be generated by any ordered set of productions.[1]

Chomsky only ever cited one paper of Post's, an informal paper on r. e. sets of positive integers that Post delivered as a lecture to the American Mathematical Society [29]. In [7] (p. 137n) and [8] (p. 7) this paper is cited as the source of the term 'generate'. Post is also acknowledged (though without a bibliographical citation) in connection with the form of Type 0 rewriting rules ([9]: 539), and is mentioned once in *Aspects of the Theory of Syntax* ([10]: 9): 'The term "generate" is familiar in the sense intended here in logic, particularly in Post's theory of combinatorial systems'. But Chomsky appears never to have made a bibliographical reference to any of Post's technical papers on production systems.[2] *SS*, perhaps because its aim was to present transformational generative grammar to undergraduate science and engineering students, has even less referencing: the bibliography includes neither Rosenbloom's book [34] nor anything by Post.[3]

3 The supposed proof that English is not finite-state

It is very widely believed that *SS* gives a proof that English is not finite-state. This is not true. A few informal suggestions are made to support the assertion that 'English is not a finite state language' so that '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). But there was no proof; and it is not clear that a proof anything like the one Chomsky seems to have had in mind can succeed.

Chomsky had given a fuller argument that natural languages are not finite-state in a celebrated technical paper of the year before: [5], cited in *SS* on p. 22. This is claimed to contain the 'rigorous proof' to which *SS* alludes on p. 23. But

[1] This is different from saying that ordering cannot restrict what a particular set of rules can generate. Pelletier [27] shows that requiring strict ordering of a set of rules can indeed make some outputs impossible to generate by that set of rules. But as he stresses, this result presumes that the set of rules is fixed, which is not the situation linguists ever find themselves in.

[2] Urquhart [41] suggests that this might be because his understanding of Post systems came from a secondary source, namely Rosenbloom [34], which Chomsky cites in [4] and [5].

[3] The contributions of Zellig Harris are also somewhat downplayed in *SS*. See Seuren [37] for discussion of the way Harris introduced top-down generation — the idea that 'a deductive system with axiomatically defined initial elements and with theorems concerning the relations among them' could be used to 'enable anyone to synthesize or predict utterances in the language.'

if the 1956 argument is sound, no one (to my knowledge) has confirmed that. I do not understand it, and nor did Daly [13]. In its original form (Chomsky [5]) it depended on a cumbersomely defined relation of "(i, j)-dependency' holding between a string S of length n, two integers i and j such that $1 \leq i < j \leq n$, and a language L over a vocabulary A. The definitions are changed in the 1965 reprint version of the paper (a footnote credits E. Assmuss for pointing out an error). The 1965 revision relies on a cumbersomely defined ternary relation of 'm-dependency' between a sentence S, an integer m, and a stringset L, where $S = x_1\, a_1\, x_2\, a_2\, \ldots\, x_m\, a_m\, z\, b_1\, y_1\, b_2\, y_2\, \ldots\, b_m\, y_m$, and there is a unique permutation of the numbers $(1, \ldots, m)$ — a bijective mapping α from $\{1, \ldots, m\}$ to itself — meeting the following condition (I quote from p. 108 of the reprint):

> "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 $c_{m+\alpha(i_j)}$ for $b_{\alpha(i_j)}$ in S_1, for each $j \leq p$."

The idea is that if in the string S the symbol a_i is replaced by the symbol c_i, restoring grammaticality in L necessitates replacing $b_{\alpha(i)}$ by $c_{m+\alpha(i)}$.

From there, the crucially relevant mathematical step is to claim that an FSL can only exhibit m-dependencies up to some finite upper bound on m (Chomsky says an m-dependency needs at least 2^m states; Svenonius [39] says this is untrue, and m states will suffice). The empirical claim is that English has no such upper bound, and is therefore not an FSL. But Chomsky does not complete the argument by connecting these abstractions to English data; he merely points to some sentence templates ("If S_1, then S_2"; "Either S_3, or S_4"; "The man who said that S_5, is arriving today" [comma in original]), and asserts that through them "we arrive at subparts of English with ... mirror image properties" and thus "we can prove the literal inapplicability of this model" (Chomsky [5], 1965 reprinting, p. 109).

Daly [13] spends many pages attempting to work out how a sound argument for Chomsky's conclusion might be based on the data that he cites. Chomsky seems to think that pairs like ⟨*if*, *then*⟩ and ⟨*either*, *or*⟩ give rise to m-dependencies. Daly could not see how this could be true. Nor can I. The words in these pairs can occur in sentences without the other member of the pair. (The same is true of other pairs such as ⟨*neither*, *nor*⟩ and ⟨*both*, *and*⟩.) It is not clear that there is *any* pair of lexical items σ and τ in English such that if $\varphi\sigma\psi$ is grammatical then $\psi = \psi_1\, \tau\, \psi_2$ with $|\psi_1| > 0$.

In addition, the reference to finding "various kinds of non-finite state models within English" (*SS*: 22–23) and the similar remark about "subparts of English with ... mirror image properties" (Chomsky [5], 1965 reprinting, p. 109) suggest a failure to appreciate that FSLs (or context-free stringsets) can have infinite non-FSL (or non-context-free) subsets. Only if such a subset can be extracted by some regularity-preserving language-theoretic operation like homomorphism or intersection with a regular set does it entail anything about the language as a whole.

Thus it is not at all clear that Chomsky ever had an argument against English being an FSL. Certainly none appears *SS*.

4 Justifying transformations

Even if *SS* had shown that natural languages were not finite-state, that would not be sufficient to justify the transformational analyses that are thought of as the book's most significant contribution, because context-free phrase structure grammars (CF-PSGs) might have sufficed. It has since been shown to most linguists' satisfaction that natural languages are non-CF (see e.g. Shieber [38]), but there was no hint in *SS* of any such result. Instead, *SS* gives three arguments based on descriptive elegance. They hinge on coordination, auxiliaries, and passives. On re-examination, all three arguments look decidedly unconvincing.

4.1 Coordination

Coordination in English is claimed in *SS* to be governed by a principle informally stated as follows ((26) in *SS*, p. 36):

(3) "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* $+ Y \ldots$)."

This is not, of course, a transformation. S_1 and S_2 are required to be 'grammatical sentences'; i.e., strings generated by the grammar. So (3) is quantifying over the entire content of the language. It is what would later be called a transderivational constraint.

The claim is not true of English. There are many cases of X and Y such that both can occur in a given context but the coordination X *and* Y cannot. Perhaps the most obvious is the case of verb agreement controllers. Let $X =$ *Don* and $Y =$ *Phil*. Then for *I think* X *was there* and *I think* Y *was there*, (3) says that *I think* X *and* Y *was there* = **I think Don and Phil was there* should be grammatical, but this is not so. Several other such failures of (3) have been noted by Huddleston & Pullum ([19], pp. 1323–1326).

Chomsky recognizes that 'additional qualification is necessary', but nonetheless claims that 'the grammar is enormously simplified if we set up constituents in such a way that [(3)] holds even approximately (*SS*, 37). In the summary rules at the end of the book (p. 113) he therefore gives a 'generalized transformation' — basically a production with two premises — to capture the effects of (3). His rule statement is given in (4).

(4) 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: $(X_1 - X_2 - X_3; X_4 - X_5 - X_6) \rightarrow$
$X_1 - X_2 +$ *and* $+ X_5 - X_3$

Remarkably, despite all the symbols, (4) is less explicit and less accurate than (3). The letter S in the variable names 'S_1' and 'S_2' might suggest 'Sentence', but S_1 and S_2 will not in fact be sentences (strings over the terminal vocabulary); they will be sentential forms (possible stages in a derivation, potentially including nonterminals). X is stipulated to be a 'minimal element', but this term is undefined—it appears to mean 'single nonterminal'. Z and W are stipulated to be 'segments of terminal strings', so S_1 and S_2 are the same string and there was no point in distinguishing them.

A case to which (4) can apply will be something like $S_1 = S_2 =$ *Put NP in the truck*. But nowhere in (4) is it guaranteed that there is any difference between the terminal strings of the X constituents in S_1 and S_2: (4) yields **Put it and it in the truck* as an output, which is probably unintended (since in (3) it was stated that 'S_1 differs from S_2').

We can assume that Chomsky intended S_1 and S_2 to be identical sentential forms that are somehow guaranteed to have distinct generated terminal strings. But nothing hangs on S_1 and S_2 at all: no use is made of the variables Z and W in the 'structural change' (the output or conclusion) of the rule. Indeed, the structural change throws away all the variables of the input: six new variables $X_1, \ldots, X_6$ are introduced, the X in the variable names have no relation to the prior uses of X. *SS* says nothing about what the X_i range over, and no connection is made between them and Z or X or W. We are left to guess that all the X_i range over terminal strings; that $X_1 = X_4 = Z$; that $X_3 = X_6 = W$; that $X_2 \neq X_4$; and that X_2 and X_4 are terminal strings of instances of the category X. None of this is made explicit in (4) or elsewhere. Nine variables are 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 an inexpert and somewhat pointless deployment of pseudo-mathematical symbolism. The content of the rule appears to be specifiable much more simply. All the rule does is to ensure that a nonterminal symbol X can exhaustively dominate the string 'X *and* X', in any context whatsoever. And a simple phrase structure rule '$X \rightarrow X$ *and* X' could have done that.[4]

Nothing is said in *SS* about multiple coordination. An attempt is made to provide for the generation of sentences like *I like indigo and violet*, but not of sentences like *I like red, orange, yellow, green, blue, indigo, and violet*. It is not made clear whether a generalized transformation can reapply to its own output, nor why $n - 2$ of the coordinators disappear in an n-coordinate structure, nor why the coordinator *and* can be placed only before the last coordinate, nor how other coordinators are introduced.

To summarize, the proposal that *SS* makes about handling coordination is obscure, incomplete, inadequate, and apparently unnecessary.

[4] It may be that Chomsky ruled out positing such a rule on the grounds that it would not permit the unambiguous reconstruction of a tree from each phrase structure derivation (see McCawley [23] on this point). But as McCawley noted, the background assumption (that trees must be built from derivations rather than licensed by phrase structure rules directly) is a strange and unmotivated one.

4.2 Auxiliaries

The *SS* analysis of the English auxiliaries is frequently cited as a novel and impressive achievement. It looks somewhat less novel when we consider the analysis published by Fries [15] five years before:

(5)

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

Fries's 'classes' are lexical categories like noun (class 1) and verb (class 2), and the 'groups' cover syntactically associated minor items like determiners (group A) and auxiliaries (group B). Fries takes the maximal auxiliary cluster to consist of a modal such as *may* followed by the perfect auxiliary *have* followed by an instance of *have to* followed by the progressive auxiliary *be*, each being optional. And the famous CF-PSG rule (6) of *SS* follows it, except that it correctly drops *have to* (not an auxiliary element at all):

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

'C' is a tense or concord (agreement) morpheme, and 'M' stands for 'modal'. So the rule lays out the tense or concord morpheme, an optional modal, an optional instance of *have* accompanied by the past participle suffix -$(e)n$, and an optional instance of *be* accompanied by the gerund-participle suffix -*ing*, strictly in that order.

Chomsky accepts Fries's idea of treating the components of the auxiliary cluster as non-verb dependents. Both defend variants of what [19] calls the **dependent-auxiliary** analysis. Fries is not explicit about how the successive items get their inflectional properties, but *SS* provides an answer: there is a transformation in *SS* (subsequently known as 'Affix Hopping') called the Auxiliary Transformation, and it is formulated thus:

(7) *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 use of symbols in the *SS* analysis is promiscuous and occasionally misleading. For example, *SS* uses no less than 6 competing and inconsistently defined symbols that might be said correspond to the informal notion 'verb': *Verb*, V, v, V_1, V_a, and V_2. The text contradicts itself about several of them. *Verb* is introduced as a lexical node on p. 28, but is clearly treated as a phrasal node on p. 39. V is introduced as a lexical node on p. 39, and is equated with the

informal term 'verb' on p. 42, but then becomes a phrasal node on p.79 (where *consider a fool* is analyzed as a V). The symbol v is an informally introduced abbreviation covering 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, yet it is mentioned in a transformation. And V_2 on p. 112 appears to stand for a subcategory of verbs (including *consider*) for which V_a was used pp. 76-77.

The text is similarly inconsistent about *Aux*. It is referred to as the 'auxiliary phrase' on p. 42, suggesting that it is a phrasal node; but on the next page it is called the 'auxiliary verb', suggesting it is a subcategory of the lexical category of verbs. This is crucially misleading, because what *SS* actually attempts to do is to analyze the syntax of English auxiliary verbs without making any reference to the notion 'auxiliary verb' at all. Nothing in the *SS* analysis corresponds to 'auxiliary verb', i.e., lexical item with verbal morphology capable of preceding the subject NP in closed interrogatives. *Aux* certainly does not correspond to that. In fact it is a very odd constituent indeed: a branching node housing a cluster of up to half a dozen non-verb siblings none of which is a head, which no transformation ever applies to or uses as a context. *Aux* is never moved, deleted, copied, inserted, targeted by adjunction, or mentioned as the context for the application of some other rule.

How or why the *SS* analysis of auxiliaries came to be regarded as elegant or attractive is not clear. The analysis certainly appears to have a host of quite serious problems, such as various ordering paradoxes. Some of the problems only emerge given later advances in syntactic theory, but many are not anachronistic in this way, and should have been apparent at the time. The most serious of these is that the analysis is simply not compatible with formal theory of Chomsky's magnum opus *The Logical Structure of Linguistic Theory* [4]. As noted by Sampson [36], the Auxiliary Transformation is not a legal transformation at all under the theory of *LSLT*. The reason is the cover symbols v and *Af*. These are neither terminal symbols nor non-terminal symbols; they function merely to make possible a collapsing of 16 different transformations sharing most of their structure.

A less abstract but still theoretical issue is that the grammar proposed in *SS* assigns such different phrase structures to sequences that we would expect to have very similar structures: *is asleep* has *is* as a V but *is sleeping* has it as a member of the Aux sequence; *ought to have left* would apparently be monoclausal but *thought to have left* is biclausal; in *has control* the word *has* is a V but in *has controlled* it is not; and so on. The arbitrary syntactic distinctions drawn have no motivation.

The fact is that modern analyses have without exception abandoned the *Aux* node. All of the items formerly housed in *Aux* are now treated as heads of projections, 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 writers as Ross [35], McCawley [24, 25], Newmeyer [26], Pullum & Wilson [32], Gazdar et al. [16], Huddleston [17, 18], and many others.

The specific version adopted by Huddleston & Pullum ([19], Ch 14, §4.2, pp. 1209ff) shares with accounts like those of Pullum & Wilson [32] or Gazdar et al. [16] a treatment of the auxiliaries of English as verbs that have certain special behaviors but take complements in the same way that other complements do. More specifically, *The Cambridge Grammar* [19] 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 (*may seem to want to avoid appearing to have been* . . . , etc.). It is now well known that VP ellipsis phenomena, negation facts, and many other considerations argue for a uniformly right-branching structure of this kind.

All in all, the treatment of auxiliaries in *SS* can hardly be said to be a progressive movement in syntactic theory or a good advertisement for transformations.

4.3 Passives

The analysis of passive clauses in *SS* is motivated by reference to four alleged problems that arise if passives are treated with phrase structure rules. According to Chomsky these complications ensue:

1. When *Verb* is expanded as *Aux* − *V*, the element *be* + *en* can be selected under *Aux* only if the *V* is transitive, and stating this would complicate the rule system (a child of *Aux* is dependent on features of a sibling of the parent of *Aux*).
2. Even if *V* is transitive, *be* + *en* cannot be selected if *V* is followed by *NP*, and stating this condition further complicates the grammar (a child of *Aux* is disallowed if *NP* occurs as a sibling of its parent's parent).
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 (a child of *Aux* becomes obligatory given a certain sibling of its parent's parent).
4. Selection restrictions reverse: acceptable subject *NP*s for passive clauses will be precisely those that would be acceptable as the object in the corresponding active, and acceptable *by*-phrase objects will be precisely those *NP*s that would have been acceptable as subjects in active clauses.

The trouble is that all four of these claims are spurious.

Claim 1: Be + en with intransitives — Not all verbs occurring with *be* and a past participle are transitive:

(8) *Man is descended from apes.* (↚ **Someone descended man from apes*).
(9) *Charles is said to be gay.* (↚ **Somebody says Charles to be gay*).
(10) *Antarctica is uninhabited by man* (↚ **Man uninhabited Antarctica*).

Claim 2: Be + en with following NP — There can be an NP after the verb in a clause with *be* + *en*:

(11) *I've often been called an idiot.*
(12) *He was denied all his legal rights.*
(13) *We were shown several nice apartments.*

Claim 3: By-phrase without be + en — A passive *by*-phrase complement can occur with no *be* auxiliary:

(14) a. *We had this* [*done by an expert*].
b. *He went and got himself* [*stung by a wasp*].
c. *This car wants* [*cleaning*].
d. *The book needs* [*revising by an experienced editor*].

Claim 4: Selection restriction reversal — It has been clear since McCawley's classic paper of 1968 [23] that selection restriction issues have no place in syntax. *SS* assumed that English syntax should distinguish *John plays golf* from *Golf plays John* (the latter is referred to as a 'non-sentence'). This cannot be right. As McCawley pointed out, every semantic property of noun phrases is capable of being relevant to such putative restrictions: the property of denoting a crustacean (objects of the verb *devein*); the property of denoting a matrix (for objects of the verb *diagonalize*); and so on.

I would say that selection restrictions do not belong in linguistics, but rather in metaphysics. Which noun phrases can fill the blank in *The* ____ *thinks it is Tuesday* or other sentences with the verb *think*? Would *baby* be appropriate? What about *foetus*? *Crocodile*? *Cockroach*? *Computer*? One can readily imagine philosophical debate about the right cutoff point. Neurologists, philosophers of mind, and animal rights advocates might not agree. Turing's famous 1950 paper in *Mind* set off controversy about whether machines can think; but surely that issue is not to be settled by syntax! This fourth point of Chomsky's is clearly just a conceptual mistake. And the other three are entirely unpersuasive for syntactic reasons.

4.4 Analyzing passives

The right analysis of auxiliaries in English leads us toward an acceptable analysis of passives too. Auxiliary verbs take non-finite, subjectless, recursively nestable complement clauses with specified inflectional features. Various matrix-clause verbs take passive clauses: *be* (*was examined*), intransitive *get* (*got arrested*, transitive *get* (*got myself appointed*), *go* (*went unnnoticed*), *have* (*have someone collected*), and so on.

We are in fact dealing with two dozen distinct constructions. Passive clauses such as *liked by his classmates* or *beaten down by her troubles* or *irritated by his kids* are best regarded as 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*).

(15) a. *He was well liked by his classmates.* [passive VP]
b. *He was decidedly popular with his classmates.* [AdjP]
(16) a. *She looked beaten down by her troubles.* [passive VP]
b. *She looked weary from her troubles.* [AdjP]
(17) a. *I often got irritated by his kids.* [passive VP]
b. *I often got angry with his kids.* [AdjP]

The verbs may be in past-participal or gerund-participal inflected form (the 'concealed passive', as in *The book merits re-reading*); they may be adjectival (as with the ones taking *un-*) or verbal. And cross-cutting these distinctions are the lines dividing prepositional passives (with stranded prepositions, as in *was looked at*) from the ordinary kind (*was seen*), and separating long passives (with the *by*-phrase complement) from short passive clauses (without it).

The full array contains 24 English passive constructions, of which the *SS* transformation handles just one: the non-concealed non-adjectival non-prepositional long passive clause as complement of the copula. This one has no special priority or importance relative to the others. If the Passive transformation expressed a true generalization (we shall see below that 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.

The key special property of passive clauses 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. This property is not tied to any of the elements present in the *SS* Passive transformation.

- it is not tied to the presence of *be* + *en*, as shown by bare passives (*Ignored by his workmates, he labored alone*);
- it is not tied to the presence of *be* + *en*, as shown by concealed passives (*She needs examining by a specialist*);
- it is not tied to the presence of an immediately postverbal NP, as shown by prepositional passives (*It has often been laughed at*);
- it is not tied to the existence of a corresponding active clause, as shown passives with verbs like *rumored* and *said* (*He is said to be interested*);
- and in fact it is not tied to clauses at all, as we see from the ambiguity of *the shooting of the hunters.*

4.5 Irregularity in the set of passives

Note also that the generalization expressed by the *SS* Passive transformation is in any case massively false. The rule says:

(18) Passive transformation
Structural analysis: $NP - Aux - V - NP$
Structural change: $X_1 - X_2 - X_3 - X_4 \rightarrow$
$X_4 - X_2 + be + en - X_3 - by + X_1$

This entails very clearly that for any *NP* immediately after any sequence of $Aux - V$, a grammatical passive will result from shifting the postverbal *NP* to subject position and the original subject into a *by*-phrase and adding *be* before the head verb and inflecting the head verb in past-participial form. But there are indefinitely many counterexamples, of many interestingly different types. Perhaps the most obvious counterexamples are strings like this:

(19) *Everyone – must – hope – things will get better.*
NP – *Aux* – *V* – *NP*
X_1 – X_2 – X_3 – X_4

From this the *SS* passive transformation (since it is blind to embedded clause boundaries) will generate the ungrammatical string in (20).

(20) **Things are hoped will get better by everyone.*

Such trans-clausal cases were treated by Chomsky in [11] as a research problem to be solved by positing a constraint on transformational movement that is violated by any movement of an NP out of a tensed domain. But Chomsky's proposals fail fairly decisively (see Bach & Horn [2], esp. 284–289).

Over and above this class of examples, there are numerous lexical and semantic limitations on passivization. Bach [1] gives a significant number. Postal [31] catalogs many more. They include cases with predicative complement NPs (*Mike seemed a nice enough guy* $\not\Rightarrow$ **A nice enough guy was seemed by Mike*); measure NPs (*The fish weighed twelve pounds* $\not\Rightarrow$ **Twelve pounds were weighed by the fish*; (*This matters a lot to me* $\not\Rightarrow$ **A lot is mattered by this to me*; manner of speaking verbs (*The old man growled some bitter comments* $\not\Rightarrow$ **Some bitter comments were growled by the old man*); and many other idiosyncratic cases (*The train departed the station at dawn* $\not\Rightarrow$ **The station was departed at dawn by the train*; *George had several homes* $\not\Rightarrow$ **Several homes were had by George*; *Fred lacks finesse* $\not\Rightarrow$ **Finesse is lacked by Fred*; etc.).

The rich array of unpassivizable $NP - Aux - V - NP$ sequences tells us much about the sensitivity of passive constructions to lexical factors. The notion that it represents some kind of simple, automatic, regular, syntactic modification process, which is the central claim presented in *SS*, has no plausibility whatsoever, and provides no motivation for transformations.

5 Conclusions

Why care about a retrospective evaluation of a monograph over 50 years old? Because myths about scientific breakthroughs and results can warp perceptions

of the history of a field. Creation myths attributing everything to one individual are known in other fields too.

The truth about science is that discoveries and innovations develop over time and build on earlier developments in the field or in adjacent fields, and myths of monogenesis and individual glorification damage contemporary theorizing in at least two ways. First, they encourage scientists in the complacent maintenance of false assumptions: if almost every linguist is convinced that *SS* showed transformations to be necessary back in 1957, non-transformational research will be underdeveloped or ignored (and indeed I think in general it has been over the past fifty years). Second, they promote biased and lazy citation practices — the same old references passed from paper to paper without anyone checking the sources. Both consequences are worth guarding against.

References

1. Bach, E.: In defense of passive. Linguistics and Philosophy 3, 297–341 (1980)
2. Bach, E., Horn, G.M.: Remarks on "Conditions on transformations". Linguistic Inquiry 7, 265–361 (1976)
3. Brainerd, W.S., Landweber, L.H.: Theory of Computation. John Wiley, New York (1974)
4. Chomsky, N.: The Logical Structure of Linguistic Theory. MIT Library, Cambridge, MA (1956), microfilmed; revised version of a 1955 unpublished manuscript
5. Chomsky, N.: Three models for the description of language. I.R.E. Transactions on Information Theory IT-2, 113–123 (September 1956), reprinted with substantive revisions in Luce, Bush & Galanter ([22]), 105–124
6. Chomsky, N.: Syntactic Structures. Mouton, The Hague (1957)
7. Chomsky, N.: On certain formal properties of grammars. Information and Control 2, 137–167 (1959), reprinted in Luce, Bush & Galanter [22], 125–155; citation to original is incorrect
8. Chomsky, N.: On the notion 'rule of grammar'. In: Proceedings of the Twelfth Symposium in Applied Mathematics. pp. 6–24. American Mathematical Society, Providence, RI (1961), reprinted with slight revision in Jerry A. Fodor and Jerrold J. Katz (eds.), The Structure of Language: Readings in the Philosophy of Language, 155–210 (Englewood Cliffs, NJ: Prentice-Hall)
9. Chomsky, N.: Explanatory models in linguistics. In: Nagel, E., Suppes, P., Tarski, A. (eds.) Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress. pp. 528–550. Stanford University Press, Stanford, CA (1962)
10. Chomsky, N.: Aspects of the Theory of Syntax. MIT Press, Cambridge, MA (1965)
11. Chomsky, N.: Conditions on transformations. In: Anderson, S.R., Kiparsky, P. (eds.) A Festschrift for Morris Halle. Holt Rinehart and Winston, New York (1973)
12. Chomsky, N., Miller, G.A.: Introduction to the formal analysis of natural languages. In: Luce, R.D., Bush, R.R., Galanter, E. (eds.) Handbook of Mathematical Psychology, vol. II, pp. 269–321. John Wiley and Sons, New York (1963)
13. Daly, R.T.: Applications of the Mathematical Theory of Linguistics. Mouton, The Hague (1974)
14. Davis, M. (ed.): Solvability, Provability, Definability: The Collected Works of Emil L. Post. Birkhäuser, Boston (1994)

15. Fries, C.C.: The Structure of English. Harcourt Brace, New York (1952)
16. Gazdar, G., Pullum, G.K., Sag, I.A.: Auxiliaries and related phenomena in a restrictive theory of grammar. Language 58, 591–638 (1982)
17. Huddleston, R.: Further remarks on the analysis of auxiliaries as main verbs. Foundations of Language 11, 215–229 (1974)
18. Huddleston, R.: An Introduction to English Transformational Syntax. Longman, London (1976)
19. Huddleston, R., Pullum, G.K.: The Cambridge Grammar of the English Language. Cambridge University Press, Cambridge (2002)
20. Lasnik, H.: Syntactic Structures Revisited: Contemporary Lectures on Classic Transformational Theory. MIT Press, Cambridge, MA (2000)
21. Lightfoot, D.: Introduction. In Noam Chomsky, Syntactic Structures, second edition, v–xviii. Mouton de Gruyter, Berlin (2002)
22. Luce, R.D., Bush, R.R., Galanter, E. (eds.): Readings in Mathematical Psychology, vol. II. John Wiley & Sons, New York (1965)
23. McCawley, J.D.: Concerning the base component of a transformational grammar. Foundations of Language 4, 243–269 (1968), reprinted in James D. McCawley, *Grammar and Meaning*, 35–58 (New York: Academic Press; Tokyo: Taishukan, 1973)
24. McCawley, J.D.: Tense and time reference in English. In: Fillmore, C.J., Langendoen, D.T. (eds.) Studies in Linguistic Semantics, pp. 97–113. Holt, Rinehart and Winston, New York (1971)
25. McCawley, J.D.: The category status of English modals. Foundations of Language 12, 597–601 (1975)
26. Newmeyer, F.J.: English Aspectual Verbs. Mouton, The Hague (1975)
27. Pelletier, F.J.: The generative power of rule orderings in formal grammars. Linguistics 18(1/2 (227/228)), 17–72 (1980)
28. Post, E.: Formal reductions of the general combinatory decision problem. American Journal of Mathematics 65, 197–215 (1943), reprinted in Davis [14], 442–460.
29. Post, E.: Recursively enumerable sets of positive integers and their decision problems. Bulletin of the American Mathematical Society 50, 284–316 (1944), reprinted in Davis [14], 461–494.
30. Post, E.: Recursive unsolvability of a problem of Thue. Journal of Symbolic Logic 12, 1–11 (1947), reprinted in Davis [14], 503–513.
31. Postal, P.M.: Skeptical Linguistic Essays. Oxford University Press, New York (2004)
32. Pullum, G., Wilson, D.: Autonomous syntax and the analysis of auxiliaries. Language 53, 741–788 (1977)
33. Putnam, H.: Some issues in the theory of grammar. In: Jakobson, R. (ed.) Structure of Language and Its Mathematical Aspects, pp. 25–42. No. 12 in Proceedings of Symposia in Applied Mathematics, American Mathematical Society, Providence, RI (1961)
34. Rosenbloom, P.: The Elements of Mathematical Logic. Dover, New York (1950)
35. Ross, J.R.: Auxiliaries as main verbs. Studies in Philosophical Linguistics 1, 77–102 (1967)
36. Sampson, G.: What was transformational grammar? Lingua 48, 355–378 (1979), reprinted in Empirical Linguistics, Continuum, 2001
37. Seuren, P.: Concerning the roots of transformational generative grammar. Historiographia Linguistica 36(1), 97–115 (2009)
38. Shieber, S.: Evidence against the context-freeness of human language. Linguistics and Philosophy 8, 333–343 (1985)

39. Svenonius, L.: Review of 'Three models for the description of language' by Noam Chomsky. Journal of Symbolic Logic 23, 71–72 (1957)
40. Thue, A.: Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln. In: Skrifter utgit av Videnskapsselskapet i Kristiana, I. No. 10 in Matematisk-naturvidenskabelig klasse, Norske Videnskaps-Akademi, Oslo (1914)
41. Urquhart, A.: Emil Post. In: Gabbay, D.M., Woods, J. (eds.) Handbook of the History of Logic, Volume 5: Logic from Russell to Church, pp. 617–666. North-Holland, Amsterdam (2009)