The mental status of derivational operations

Tim Hunter — *Department of Linguistics, UCLA*

April 15, 2017

Two kinds of theories of natural language syntax can be distinguished: representational theories and derivational theories. A representational theory posits some set of constraints, and defines a well-formed syntactic object to be one that satisfies all of the constraints. A derivational theory instead takes the form of a nondeterministic mechanical procedure, for example a symbol-rewriting procedure or a procedure that builds larger objects out of smaller ones, and defines a well-formed syntactic object to be one that is generated by this procedure.

The mentalistic claims of a representational theory are relatively clear: it is generally understood that when a speaker comprehends or produces a sentence, a representational theory predicts that a corresponding well-formed syntactic object (say, a tree structure with the sentence's words at its leaves) is grasped in the speaker's mind. With a representational theory, nothing is said about how a speaker might go about constructing (a representation of) this syntactic object, and the linguist's everyday use of the theory also does not involve any descriptions of procedures that construct syntactic objects.

The situation for a derivational theory, however, is slightly less straightforward. Consider for example the mainstream contemporary derivational theories deriving from Chomsky (1995) and subsequent work. It is natural to assume that a speaker grasps the syntactic object that is the end product of the derivational process corresponding to the sentence being comprehended/produced, i.e. the tree structures that are routinely used to illustrate proposals in this literature. But if that is the extent of a derivational theory's mental commitments, what is the scientific role of the derivational procedure? If we have an existing derivational theory T_1, and an alternative theory T_2 proposes a derivational procedure that differs from that of T_1 but yields the same set of well-formed syntactic objects, then is there any clear sense in which we should understand the two theories to be different? If they are not different — i.e. if the procedural component of a derivational theory does not contribute to its empirical bottom line — then why bother with the derivational procedures at all? If they are different, then *how* are they different, i.e. *how* does the procedural component of a derivational theory contribute to the theory's empirical bottom line?

My main goal here is to propose an answer to this last question. I will lay out a way of understanding derivational theories according to which the derivational process itself, in addition to the end result of this process, plays a part in determining the empirical predictions of a theory. The important idea is to conceive of a derivational *process* as a first-class theoretical object which can underpin empirical predictions just as naturally as the static objects in a representational theory can; we can identify an atemporal structured object — typically, a *derivation tree* — that encapsulates the derivational process and yet is static in the same sense that syntactic objects in representational theories are.

The conception of derivations that I propose seems somewhat unnatural in the context of current theories, because of an unfortunate quirk in modern generative syntax: the fact that the end product of a derivation very often encodes a large amount of, or even all of, the derivational process itself, for example in the form of co-indexed traces or copies. A consequence of this is that very often the derived syntactic objects themselves are all one needs to look at in order to

determine a theory's predictions about acceptability judgements; any number of different ways of describing how those objects are constructed would seem to leave a theory's empirical footprint intact. This has led some to express skepticism about the empirical role of derivational operations in modern generative syntax (Sag and Wasow, 2011; Jackendoff, 2011; Ferreira, 2005; Phillips and Lewis, 2013).

I have two points to make in response to this line of thinking.

First, if it is agreed that we might hope for our theories to make predictions in domains beyond acceptability judgements, then we should not set aside derivational operations simply because they have no impact on acceptability predictions. It may be that the derivational differences between two otherwise-equivalent theories can be linked to differing predictions about, say, comprehension difficulty, or some other "finer-grained" properties of sentences. This of course raises the question of exactly what linking hypotheses might be used to expose derivational operations to the empirical spotlight. Spelling out such linking hypotheses is a goal of my ongoing work; for reasons of space, here I can only "set the stage" for the overall argument. But importantly, there do exist hypotheses that link derivational processes to, for example, comprehension difficulty *without* incorporating the hypothesis that derivational operations are real-time operations carried out in the course of producing or understanding a sentence.

Particularly given the "quirk" mentioned above, it may seem that my method of giving derivations empirical bite is an unnatural post-hoc reinterpretation of what a grammatical derivation is. The second point I want to make is that appealing to the derivation in its entirety as the source of empirical predictions is in fact entirely unremarkable when we consider the overall history of generative grammar: the derivational details themselves have often had an important role to play *even in the prediction of acceptability judgements*. So the current quirk that makes derivational processes seem inconsequential is an anomaly, given the big picture.

A clearer understanding of how derivational grammars in general can constitute hypotheses about mental phenomena can be achieved by considering other derivational systems that do not have this quirk, and where it is therefore easy to see the role of the derivational process itself (because this role is not duplicated by representational devices).

Purely representational and purely derivational systems

Caricaturing at least slightly, Figure 1 illustrates one possible conception of the relationship between a representational system and a derivational system. On the left is the static syntactic structure assigned to the sentence 'Kim gives Sandy Fido' in HPSG, one of the more widely-known representational theories of grammar (Pollard and Sag, 1994, p.33). This syntactic object is well-formed by virtue of satisfying the relevant array of constraints. As mentioned above, the mental commitments of this kind of theory are relatively clear: in comprehending or producing this sentence, a representation of this syntactic object is grasped[1] in the speaker's mind. By virtue of the fact that this grasped syntactic object is well-formed, the theory predicts that this sentence will be judged to be acceptable. And perhaps there are other predictions that one might make on the basis of other properties of this syntactic object: to take an overly simplistic example, one might predict that the time taken to comprehend this sentence will be some function of the size of this object.

On the right of Figure 1, for comparison, is a sketch of how a derivation on modern minimalist grammar might be thought of. There is a final derived expression of the familiar sort, the tree with yield 'the dog will chase it' shown at the top. One possible thought — although I will

[1]I will assume that the intended meaning of this term, while difficult to spell out explicitly, is sufficiently clear. Since the questions I aim to address here largely centre on the difficulties that come with adopting derivational as opposed to representational grammars, I am taking as my concrete goal to show that there are no such *additional* difficulties. Fleshing out the notion that I am calling "grasping" is a difficulty that will affect derivational and representational theories equally.

argue against this — is that *this* tree is the thing in this derivational system that best corresponds, as indicated by the horizontal dashed lines, to HPSG's static syntactic object on the left. Since this is a derivational framework, however, there is more to the picture than just this: there is also a derivational process which is taken to have given rise to this derived expression, as shown underneath. There is nothing in the illustration of the representational system on the left which corresponds to this derivational process. So what is it there for?

I will argue that instead of this view, we should consider the derivational process as a whole (including, but not limited to, the final derived expression) to be the analog of the static representation in a representational system. This shift in perspective is reflected in the shift from Figure 1 to Figure 2. The arrows that are usually thought of (and can still be, harmlessly) as indicating a kind of precedence are now simply part of the object that a speaker must grasp; the formal *relationships amongst expressions* that they express are part of the information that a speaker must recover.

It will be useful to establish some terminology for what follows. I will use the term *expression* for an object of the sort that might be manipulated or inspected by a grammar: either checked for consistency with some representational constraint, or used as input to or produced as output from some derivational operation. I will show expressions inside rounded double boxes throughout. I will use *object* as a much more general term for any kind of structured representation that a mind might grasp. Expressions are objects, but not all objects are expressions. In a representational setting, there are no relevant objects to consider besides expressions themselves, and so the object to be grasped upon encountering the sentence 'Kim gives Sandy Fido' is simply the expression itself that appears on the left of Figure 1 and Figure 2. The difference between these two figures is that Figure 1 expresses a view where, in the derivational system, the object to be grasped is the single expression shown within the horizontal dashed lines; whereas Figure 2 expresses the view that the object to be grasped is an object of a different sort, an object encoding certain relations among expressions. This object is a derivation (and can be represented on paper by a derivation tree).

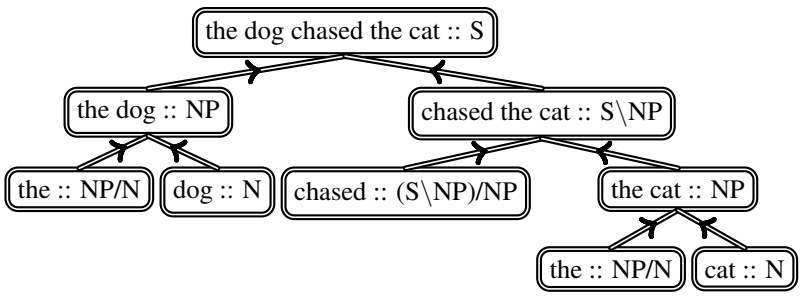
A clear illustration of the perspective presented in Figure 2 is provided by the various kinds of categorial grammar. In this framework, the categories into which lexical items are classified can be complex, and a small number of very general combinatory rules apply in a manner that is guided by these potentially complex categories. For example, using the lexical items shown in (1), the two general rules of forwards and backwards function application can be applied recursively to construct the sentence 'the dog chased the cat'. This is typically illustrated using a format like (2), but an equivalent representation that follows the conventions I adopt throughout this paper is the one in (3).

(1)
the :: NP/N
dog :: N
cat :: N
chased :: (S\NP)/NP

(2)

$$\dfrac{\dfrac{\dfrac{\text{the}}{\text{NP/N}}\quad\dfrac{\text{dog}}{\text{N}}}{\text{NP}}\qquad\dfrac{\dfrac{\text{chased}}{\text{(S\NP)/NP}}\qquad\dfrac{\dfrac{\text{the}}{\text{NP/N}}\quad\dfrac{\text{cat}}{\text{N}}}{\text{NP}}}{\text{S\NP}}}{\text{S}}$$

(3)



A distinctive feature of this kind of grammar is that the expressions being manipulated are es-

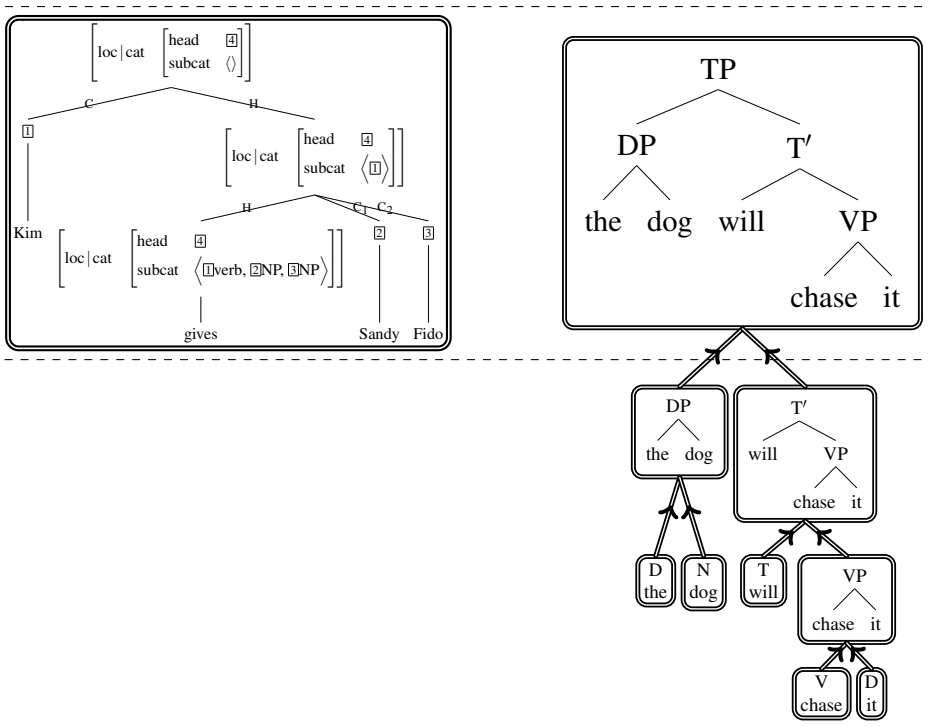


Figure 1: A view that I will argue against: only the end product of a derivational process is given the easily-understandable empirical status corresponding to that of a static representation in a non-derivational theory.

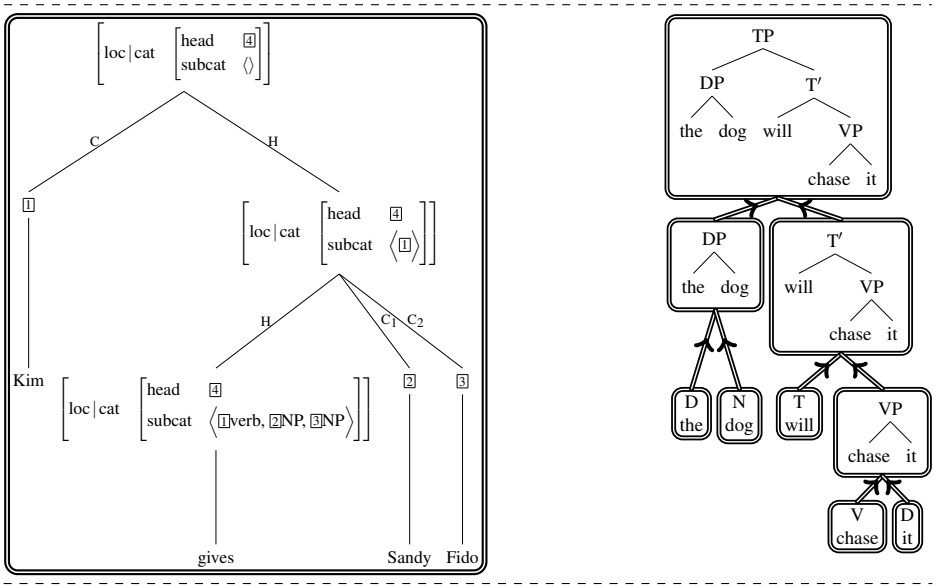


Figure 2: The view that I will argue for: the derivational process itself, in its entirety, is the relevant object.

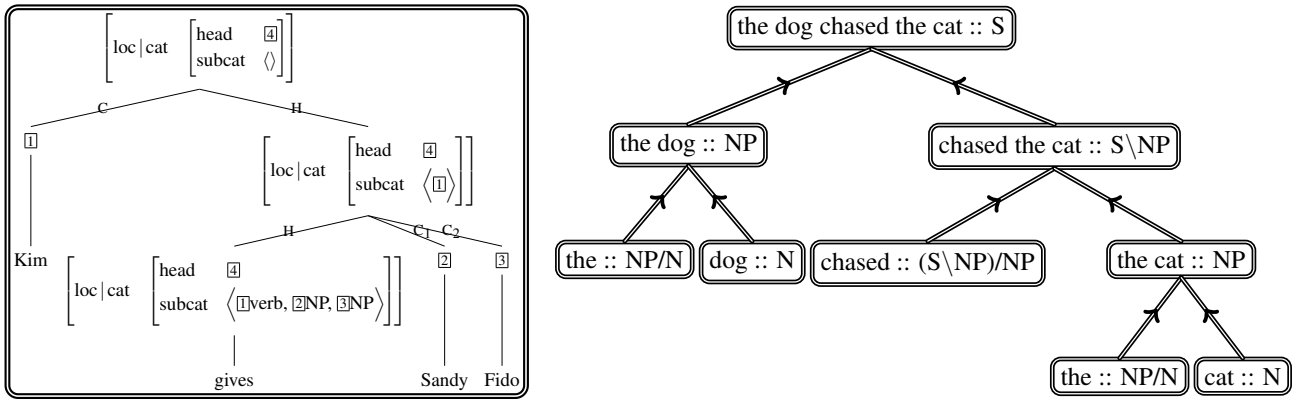


Figure 3

sentially unstructured: they are things like 'the dog :: NP', i.e. a string paired with a category, where the category dictates how the expression can be used by any subsequent operations. So the derivational process indicated in (2) and (3) is one which works with the "ingredients" shown in (1), and produces as a result the expression 'the dog chased the cat :: S'. Notice that the final derived expression is an object of the same sort as the ingredient expressions in (1), i.e. a string with a category. As Jacobson (2007) describes this kind of system, "there is no room to state constraints on structured representations. For 'structure' is not something that the grammar ever gets to see". In the terminology introduced above, this is to say that the expressions here, the things that the grammar can "see" — inspect, manipulate, whatever — have no structure; the only structured object is the derivation.

The crucial point here is that it would make little sense to suppose that the object that is "grasped" by a speaker upon encountering the sentence 'the dog chased the cat' is simply the one *constructed by* this derivational process, namely the expression 'the dog chased the cat :: S'. For the theory to be doing any work at all, there must at the very least be some difference between what the speaker does upon encountering 'the dog chased the cat' and what he/she does upon encountering 'cat dog the the chased'. But this is not a difference between 'the dog chased the cat :: S' being well-formed and 'cat dog the the chased :: S' being ill-formed relative to some constraints on representations — there are no such constraints. Rather, the difference is that in the case of 'the dog chased the cat', *there is some derivational process* that produces the expression 'the dog chased the cat :: S', whereas in the case of 'cat dog the the chased' there is no derivational process that produces the expression 'cat dog the the chased :: S'. So what is grasped by a speaker encountering 'the dog chased the cat' is some representation like (3): something that encodes the relationships between the ingredients like 'the :: NP/N' and 'dog :: N' and the things that are built from them like 'the dog :: NP'. It is clear, then, that in this kind of system the derivational process is doing some real work, in such a way that it makes sense to construe the derivational process itself as the object that corresponds to the representations to be grasped in the setting of a representational theory; see Figure 3.

What makes the importance of the derivation so clear in categorial grammars is the fact that, as emphasized above, the expressions constructed by these derivations are just strings (with categories) that have no significant structure. Thus there is, roughly speaking, "nothing but the derivation", and so when it comes to asking what the theory says about (what a speaker will do upon encountering) a particular sentence, the derivation itself is the only thing to look to. But the general point can be carried over to systems where the derived expressions have more structure, for example, if they are trees rather than strings: in such systems, it is less obvious that it is *necessary* to treat the derivation with the significance indicated in Figure 2 and Figure 3, but there is no obstacle to doing so if it is useful.

Mixed systems

Whereas HPSG is a purely representational system and categorial grammar is a purely derivational system, transformational grammar is what I will call a mixed system: a transformational grammar incorporates a derivational component (like categorial grammar), but the expressions constructed are structured (unlike categorial grammar) and these structures are subject to certain representational constraints as well (like HPSG). The quirk mentioned in the introduction makes modern generative grammars a system that is mixed but largely *redundant*: very often, what the grammar enforces via a derivational constraint could just as easily be enforced via a representational constraint, and vice-versa. But I argue that this impression is misleading.

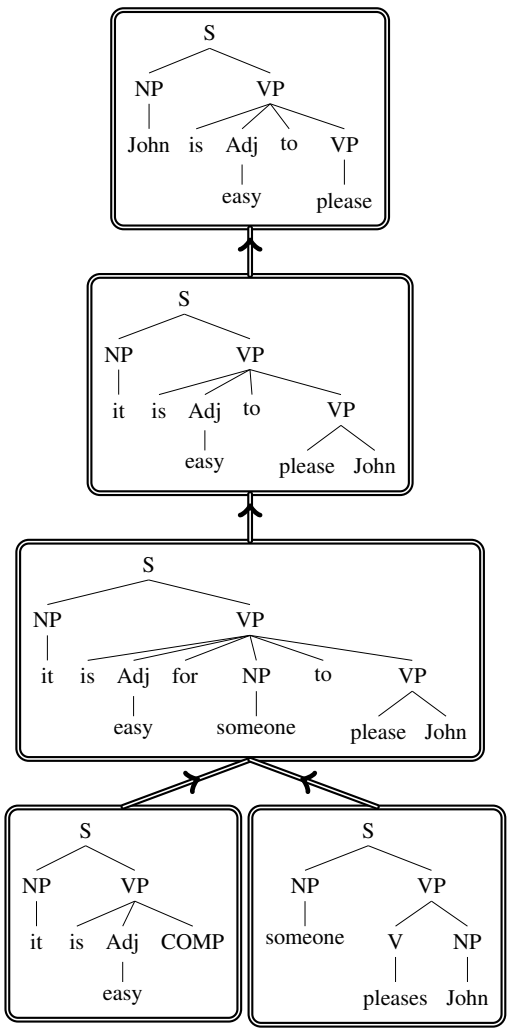
One example of a mixed but clearly non-redundant system is the framework of early transformational grammars in Miller and Chomsky (1963) and Chomsky (1965). A clear illustration of this comes from the famous comparison between the two sentences in (4) and (5) (see Miller and Chomsky, 1963, pp.476–480).

(4) John is easy to please.

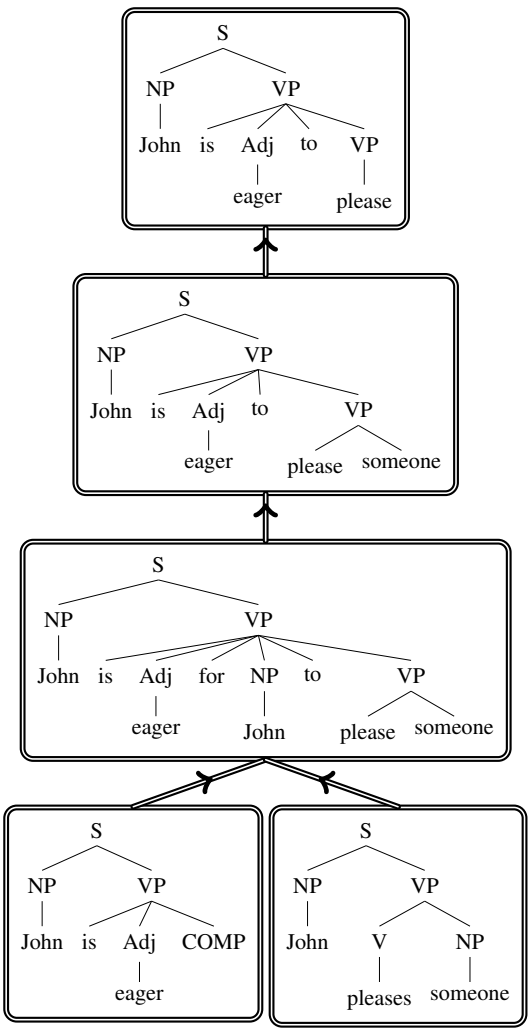
(5) John is eager to please.

Each of these sentences is derived by base-generating two monoclausal "underlying P-markers", and then manipulating and combining these P-markers (these are the expressions that this system works with) to arrive at a single "derived P-marker", as illustrated in (6) and (7).

(6)



(7)



This is a derivational system that works with structured expressions (specifically, trees, rather

than strings), so this is a mixed system. The grammar licenses certain derivational steps that relate P-markers to one another — specifically, certain transformations, such as the transformation that combines two S-rooted trees and the transformation that fronts an NP from an embedded object position to overwrite 'it' — and also imposes certain representational constraints ("surface filters"[2]) on the eventual derived expression. But the work that is performed derivationally and the work that is performed representationally are separate, and accordingly grasping the entire derivational process provides more information than does grasping the final derived expression alone.

Furthermore, it is clear that the intended interpretation of these early transformational grammars *did* involve the idea that a speaker encountering (4) or (5) grasped the entire derivational process illustrated in (6) or (7). This pair of sentences provides a dramatic illustration of this: the interesting point about this pair is that speakers understand them to have different structures in some important sense — as evidenced by the fact that speakers understand 'John' to be the pleasee in (4) but the pleaser in (5), and the fact that speakers know there is an expletive-'it' variant of (4) but not (5), etc. The crucial point to note is that the theory would not provide any account of these differences if one supposed that the object grasped by speakers were simply the eventual derived structures, because these two structures are identical (modulo the alternation of 'easy'/'eager' itself), as (6) and (7) make clear. In order to provide any explanation for the different ways in which speakers treat these two sentences, the derivational processes posited by the theory, i.e. the entire tree structures shown in (6) and (7), must be the objects thought to be grasped by speakers.

This point is not only logically necessary in hindsight, but was clearly the intended interpretation at the time:

> . . . we see that the grammatical relations of 'John' and 'please' in [(4)] and [(5)] are represented in the intuitively correct way in the structural descriptions provided by a transformational grammar. The structural description of [(4)] consists of the two underlying P-markers [at the bottom of (6)] and the derived P-marker [at the top of (6)] (as well as a record of the transformational history T_1, T_4, T_5). The structural description of [(5)] consists of the two underlying P-markers [at the bottom of (7)] and the derived P-marker [at the top of (7)] (along with the transformational history T_1, T_2, T_3). Thus the structural description of [(4)] contains the information that 'John' in [(4)] is the object of 'please' in the underlying P-marker [at the bottom right of (6)]; and the structural description of [(5)] contains the information that 'John' in [(5)] is the subject of 'please' in the underlying P-marker [at the bottom right of (7)].
>
> . . . [one component of the perceptual model] will utilize the full resources of the transformational grammar to provide a structural description, consisting of a set of P-markers and a transformational history, in which deeper grammatical relations and other structural information are represented.
>
> Miller and Chomsky (1963, pp.479–480)

This kind of transformational grammar is therefore a mixed system where the final derived expression does not provide all of the grammatically relevant information — like in (3) above. So having trees instead of strings as the derived expressions does not automatically make the derivational process redundant.

With the rise in representational devices such as traces/copies and co-indexed silent elements like PRO, cases like (6) and (7) where the derivation clearly provides additional information beyond what can be gleaned from the final derived expression have become rarer and rarer. But some more recent analogous cases can be identified, so there is no clear reason to assume that the

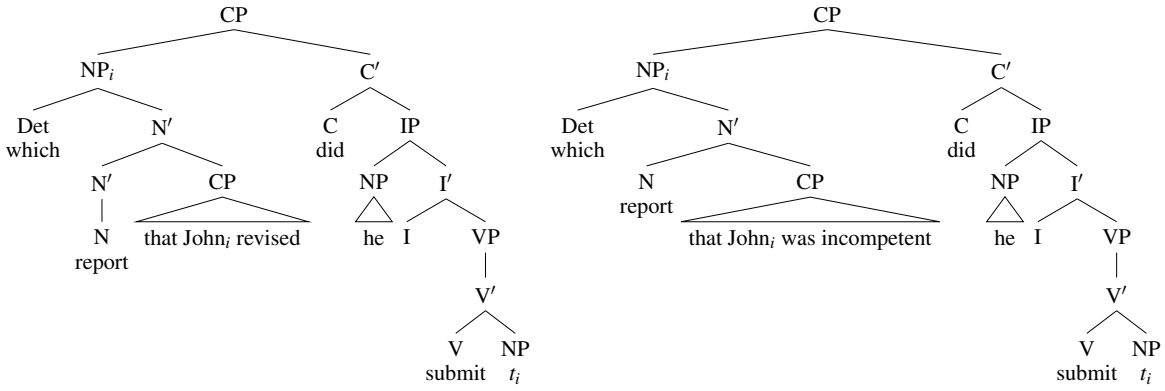
[2]This is not strictly true of the transformational grammars of the 1960s: filters as we currently know them were not introduced until the 1970s . . . but they were introduced into a system that still routinely "destroyed" information as the derivation progressed, so the crucial point remains.

derivation-centric view expressed in the quotation above has ever been deliberately discarded. To take one well-known example, consider the analysis of the contrast in (8) proposed by Lebeaux (1988).

(8) a. Which report [that John$_i$ revised] did he$_i$ submit?
b. * Which report [that John$_i$ was incompetent] did he$_i$ submit?

Lebeaux's influential account of this contrast involved supposing that the relative clause in (8a) could be added after the wh-movement transformation that fronts 'which report', since the relative clause is not required to be present in d-structure. The bracketed clause in (8b), however, being a complement rather than an adjunct, does not have this flexibility, and therefore has no way to avoid the Condition C violation induced by the co-indexed matrix subject 'he' at d-structure. The crucial point for our purposes here is that the two trees shown in (9) do not *themselves* differ in any respect that is relevant to compliance with Condition C. The theory would not provide any account for the fact that speakers' judgements of (8a) differ from their judgements of (8b) if it were assumed that speakers grasped only the trees in (9). This is analogous to the way the early transformational grammars would not provide any account for speakers' differing judgements regarding the 'easy'/'eager' contrast if it were assumed that speakers grasped only the derived P-markers shown at the top of (6) and (7). Instead, the theory must be interpreted as claiming that speakers grasp the entire derivational process.

(9)



Conclusion

The preceding discussion has had two aims. The first aim was to establish what it looks like for a theory of grammar to suppose that the objects being grasped by a speaker are derivations. This comes out most clearly in the case of systems like categorial grammar, but importantly there are also mixed systems that derive structured expressions (for example, trees) and yet also require this same derivational interpretation. The second aim was to demonstrate that although it is no longer as clearly the case as it was in the early days of the 1960s, generative grammar has never ceased to be a system of the mixed kind that takes derivations themselves to be the objects to be grasped by speakers. The instances where this can be seen have become rarer over the years for unrelated empirical reasons (which arguably are less relevant now than they were in the GB era), but certain specific well-known points in the literature make it clear that this is still the intended interpretation.

If we accept this conclusion about the cognitive commitments of modern generative grammar, then we should expect that in principle there will be ways to empirically distinguish theories on the basis of their derivational claims — not the claims they make about derived expressions (for example, which ones are well-formed and which ones have which particular interpretations), but the claims they make about the derivational processes of which those expressions are the end result. Cashing out these claims is a goal of my current research.

References

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press, Cambridge, MA.

Ferreira, F. (2005). Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review*, 22:365–380.

Jackendoff, R. (2011). What is the human language faculty?: Two views. *Language*, 87(3):586–624.

Jacobson, P. (2007). Direct compositionality and variable-free semantics. In Barker, C. and Jacobson, P., editors, *Direct Compositionality*, pages 191–236. Oxford University Press, Oxford.

Lebeaux, D. (1988). *Language acquisition and the form of the grammar*. PhD thesis, University of Massachussetts, Amherst.

Miller, G. A. and Chomsky, N. (1963). Finitary models of language users. In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Handbook of Mathematical Psychology*, volume 2. Wiley and Sons, New York.

Phillips, C. and Lewis, S. (2013). Derivational order in syntax: evidence and architectural consequences. *Studies in Linguistics*, 6:11–47.

Pollard, C. and Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Sag, I. A. and Wasow, T. (2011). Performance-compatible competence grammar. In Borsley, R. and Borjars, K., editors, *Non-Transformational Syntax: Formal and Explicit Models of Grammar*. Wiley-Blackwell.