On the Convergence Of 'Minimalist' Syntax and Categorial Grammar

Professor Robert C. Berwick, MIT Department of Computer Science

Professor Samuel David Epstein, Harvard Department of Linguistics
Cambridge, MA 02139

Abstract

This paper shows that the so-called "Minimalist Program" of Chomsky (1993, 1995) can be given a natural interpretation as a categorial system in which there is exactly one syntactic (algebraic) operation: namely, "Hierarchically Concatenate" (HC) (what Chomsky calls "Merge"), and also replacing the representations of "D-structure," 'S-structure" and transformations with the derivation lines typical of categorial systems—thus unifying two previously disparate approaches to the analysis of natural language. For example, the general "movement" rule of transformational grammar is easily seen to be a subcase of Hierarchical Concatenation of (alpha, beta), where alpha is a subtree of beta; this automatically derives the usual c-command condition on so-called "empty categories." The usual semantic interpretation benefits of the categorial approach follow directly.

Further, it demonstrates that by positing a single syntactic concatenation operation one can *derive*—rather than stipulate—the observed grammatical relations in natural languages (viz., Subject-Verb, or Specifier-Head; the notion Head-of; Verb-Complement or Head-Complement; and constituent command or c-command), as well as the primacy of "adjacency" in syntactic constraints.

Finally, this paper shows how the "minimalist program" can be extended to the computational ground of parsing, in that the concatenative system can be naturally interpreted as a generalized, canonical LR parser with a corresponding minimal set of computational operations—suggesting that the (abstract) human parsing system is, like the human language faculty, "perfect" in the sense that the parser delivers to the language faculty exactly those derivational sequences required for the language faculty to "interpret" sentences.

1 At the Cartesian Well: The Minimalist Program & Categorial Grammar

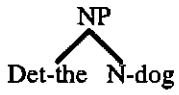
Imagine the following scene. You are at your favorite beer hall somewhere in Amsterdam—let's call it the Cartesian Well. Well known meeting place of intelligentsia, you are not surprised when a thin person dressed all in black sidles up to you and whispers in your ear, "Have I got a linguistic theory for you!" You of course yawn, have heard many such fables in your time; besides you have drunk too much. "No, wait," the figure grabs your shoulder, "I've discovered that Chomsky's latest approach to syntax and categorial grammar are *converging*."

Another flat-earth cultist? you think. "Well, hear me out—at least let me buy you another beer." So you to listen to the tale:

- Chomsky dubs his research program *Minimalist Syntax*. The idea is that you don't want to posit any syntactic entities at all beyond what's absolutely necessary for linguistic description and explanation.
- What machinery *is* necessary? Lexical items of course—but, minimally, only those, plus elements "composed from" lexical items. (More about composition in a moment!) One begins sentence generation with essentially a multiset, or *enumeration* of those items. For instance, for the sentence ultimately generated by the syntax as *The dog likes it* we would have the (unordered) enumeration {the, dog, likes, it}.
- There are no indices, subscripts, bar levels, no phrase names like NP or VP—indeed, no X-bar theory at all. (So much for the intricacies of sub- and supra- indexing in binding.)

• More generally, we dispense with the old transformational generative grammar picture of "levels of representation" such as D-structure, S-structure, LF, and PF, arranged in the familiar inverted Y-diagram with transformational rules mapping between them. Instead, there are only two (natural) representations that stand for the two (natural) interfaces between the language faculty and the rest of the mind/brain and world: one external, namely, the interface to motor systems of speech and perceptual systems of parsing; and the other, the interface to the other cognitive systems of thought, inference, and the like. We may regard this as the longstanding conventional view of language as (sound, meaning) pairs.

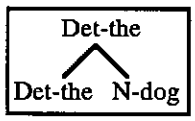
• Lexical items are built into more complex objects by a *single* compositional operation called *Merge*, more readily thought of as **Hierarchical Composition** (HC). That is, HC takes two hierarchical representations as input (these will be defined shortly), and produces a new, extended hierarchical representation as output, with one of the two inputs selected as the *head* or *root* of the extended representation. [1] For instance, following conventional notation (which we shall dispense with shortly), we would compose *the* (a Determiner) and *dog* (a Noun) as follows, projecting the Noun as the Head:



In reality, since there can be no phrase labels — these are not lexical items — HC takes as input the two lexical items `Det-the` and `Noun-dog` and composes them, selecting `Det-the` as the head or functor (the so-called DP hypothesis).[2]

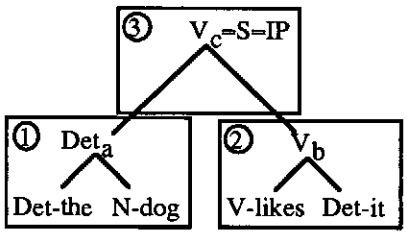
$HC(X, Y) \rightarrow \{X\{X, Y\}\}$e.g.
HC(Det-the, Noun-dog} $\rightarrow$ {Det-the, {Noun-dog}}

where the features of the functor `Det-the` have been *projected* (i.e., copied) to the Head of the composed item, the first element of the set, or its *label* (see the figure below). A **sentence derivation** thus consists of (i) initially selecting a multiset of lexical items; and then (ii) at each step, selecting a pair for input to HC consisting of a selected lexical item and another lexical item or a set resulting from a previous application of HC. In what follows, we shall often identify the composed result of HC simply by its label, or even more simply by an abbreviation for its label, e.g., "Dthe" for *the* and "Da" for the compositional result of $HC(Dthe, Ddog)$, drawing a box around the composed elements. Informally:



Pursuing the minimalist ideal, we ask why HC takes the form that it does: why does it compose only two items to yield a third? Answer: clearly, HC makes no sense operating on just *one* item. Its minimal arity is two. Because two arguments evidently suffice (empirically) for natural languages (more deeply: perhaps all grammatical relationships are expressed between *adjacent* syntactic items, that is, natural grammatical systems are *noncounting* in the technical sense of McNaughton and Papert, 1967), we do not require (at this point) higher arities for HC. (Binary branching hierarchical structure, independently motivated in current linguistic theory, follows as a result.)

To see how this all fits together so far, let us consider the derivation of the simple sentence *the dog likes it* that could be described equally well by a context-free phrase structure grammar, or a categorial grammar (example from Epstein (1995:8)). The numbered boxes denote successive applications of HC, and we have omitted much inessential syntactic detail (e.g., Inflectional morphology, etc.). It is to be stressed that this picture is entirely illustrative; the actual representations are simply the derivation lines.



Derivation lines:[3]

[1] Hierarchical composition mirrors the linear concatenation of adjacent items in phonology.

[2] See Abney, 1987 and discussion below.

[3] We have put to one side here a minor technical matter and one important issue. The technicality is that the numeration set is not actually a multiset—its duplicate members are distinguished by "some means" that we shall not cover here. The second, more important issue is the question of how a "nonbranching" lexical item like *it* or *John* can furnish a function-argument pair as required for input to HC. Roughly, we follow Chomsky's (1993) assumption that these are functors with a single additional empty argument; there are in fact some linguistic arguments for this point. This po-

Operation	Result
Select *Det-the*	{ *Det-the*}
Select *Noun-dogs*	{ *Noun-dogs* }
HC	D_a
Select *V-likes*	{ *V-likes*}
Select *Noun-it*	{ *Noun-it* }
HC	D_b

On this view, the derivation of a sentence is simply a line sequence (i.e., proof) of HC operations (lexical selection and HC merger), starting from the lexical multiset. There is no "D-structure" (i.e., representation of lexical items with their thematic roles arrayed in a hierarchical tree) nor any "S-structure". At any derivation line, the system can decide to "pronounce" — pass to the phonological or motor-articulatory apparatus — its current line derivation. If the motor component can "speak" (spell out) the derivation, then fine; if not, then that proof tree fails (e.g., the system could decide to stop, incorrectly, at the representation of *the dog likes*, and spell out only those items.

The relationship to classical categorial grammar should be apparent, putting aside for the moment the questions of semantic interpretation and so-called "phrase movement", to which we return immediately below. Of course, in some cases, the result may not be well-formed (e.g., we could decide to concatenate *the dog* and V_b first,but this would in effect make *the dog* the Object (i.e., bear the Patient thematic role); still other possibilities, like the concatenation of *Det-the* and *V-likes* yield ill-formed structures; above we show just one of the well-formed derivational sequences). Note that so far we have not yet said how the system "decides" that the output label (i.e. root) of HC is either *Det-the* or *Noun-dog*.

As in categorial grammar, the *choice* of functor is entirely a property of lexical items—in classical terms, whether we view *the* as a functor taking an NP argument to its "right" or *dog* as a functor taking a Determiner argument to its "left". Whatever choice is made here, the point is that it is an argument-taking property of a now-complex lexical item—again a familiar notion. There is one twist, however: note the deliberate "scare quotes" around the word "right" and "left." In actuality, we assume that syntactic structure (as opposed to phonological structure) expresses only hierarchical relations and not left-to-right precedence (which is a property of the external, temporal world). In other words, the only choice really made is whether *Det-the* or *Noun-dog* is selected as the functor; the directionality at the level of syntax is immaterial—a property of the phonological component, perhaps. Summarizing so far then, we have gained, first, the insight that the "Minimalist Program" is really a version of categorial grammar; and second, that the directionality in classical bi-directional categorial systems is, on this view, an artifact of not separating the "physical interface" level of temporal ordering (phonological order) from the purely syntactic operation of HC. (We shall see below that this separation of precedence from hierarchical or dominance relations has other welcome consequences, namely, it entails the possibility of ambiguous syntactic relations as in Prepositional Phrase attachments, as well as ambiguous quantifier readings.)

Minimally, then, it appears that natural language requires some operation like HC that accounts for the "is-a" (constituent) relations of language. In the (minimally) best of all possible worlds, no other operations need apply. However, it appears at first glance that natural language contains familiar "displacement" operations that move elements around, e.g., the filler-gap relations such as *What did John eat* or *John was arrested* described in a variety of grammatical frameworks via "movements" or "slashed categories" or more local operations described in still other accounts such as Bach's (1983) by "wrap" operations whereby a lexical category such as $(a/b)/c$ is flips its functor around to $(a/c)/b$, as would happen in, say, auxiliary verb movement or affix hopping.

As is well known, the "long distance" filler-gap dependencies display certain constraints: for example, a "filler" such as *Who* or *John* in *Who did you see* must *command* its "gap"—the inaudible empty argument position. (E.g., in generalized phrase structure grammar this property is ensured by only introducing slashed categories along with inaudibilia in the same right-hand side of a context-free rules. However, in the current system, we can derive this property "for free"—the first recognizable advantage in adopting the minimalist metaphor. Suppose there is in fact only *one* syntactic operation, HC. Then in fact we can *derive* the possibility of "displacement" as a special case of HC with exactly the desired property that "fillers" c-command their "gaps." Namely, take "displacement" to be that case where we have $HC(\alpha, \beta)$, and $\alpha \in \beta$ (i.e., conventionally, a subtree of β. Then this will derive

sition is plainly problematic for other reasons, among these, that the standard solutions offered by Curry (1961) among others have not been dealt with in detail; but we defer discussion of the issue here.

exactly the cases of wh-questions, topicalizations, etc.—movements generally. Thus in fact, the syntactic components needs only *one* operation.[4]

Turning next to semantic interpretation, we may regard this as the analog of "pronunciation" in the domain of "logical form" or semantics—that is, the interface to the cognitive systems of interpretation, inference, etc. Here too the minimalist program follows categorial grammar rather directly. As each "box" in the third figure above is completed, in the numbered order given—that is, as each HC operation is carried out—the resulting structure is directly (and transparently) interpreted via a "standard" Montagovian approach (or one may substitute one's own favorite semantic/intensional account here without undue strain). For reasons of space, not much more will be said here about semantic interpretation; the chief point is that the virtually 1-1 correspondence with categorial grammar makes it easy to adopt all the virtues of semantic hygiene that Bach (correctly) advocates:[5]

> "If we can take the relations between syntax and semantics as a guide, we would take a homomorphic relation to be the unmarked case, with apparent departures from it providing the most interesting challenges." (1988:32)

2 Explaining Grammatical Relations

Besides the straightforward connection to categorial grammar, minimalist syntax offers several advantages that do not seem to have been so far widely recognized. Epstein (1995) has remarked that one of the facts we must explain about language is *why* we observe only certain grammatical relations and not others. For instance, given a sentence like *the dog likes it*, one would commonly list the following as the significant (perhaps only) grammatical relations:

- *the dog* stands in the Subject-of relation to the Verb, more generally called the *Specifier-Head* relation;
- *it* and *likes* stand in the Object-of relation, more generally, the *Head-Complement* relation (or, perhaps, sister-of or govern);
- The VP (old notation) dominates *it*, etc.;
- The Subject NP *the dogs* c-commands the VP, the V, and the Object NP, but the Object NP does not c-command the Subject (where c-command is taken to be: α c-commands β iff the first branching node that dominates α also dominates β).

This is not meant to be an exhaustive list, but it does illustrate an important point: (1) Why are these relations expressed, but no others out of the infinity of possible relations among two elements in an arbitrary hierarchical structure; (2) Why do the definitions/relations have the form they do — i.e., why is c-command *stipulated* as "the *first* branching node..." rather than, say, the seventh?

Epstein's answer to this question is straightforward: HC provides the "visibility" conditions for all and only the possible grammatical relations. Take for example the relation "c-command." Given the line-by-line "proof" for sentences in the minimalist framework, we can now *derive* c-command as follows (Epstein's example 16):

(1) X c-commands all and only the terms of the category (label) Y with which X was paired by HC *in the course of a derivation.*

Thus, in our example *the dogs like it*, the Determiner (Specifier) D_a= "the dog" c-commands V_b= the VP and all terms of V_b, because it was merged with the VP during the course of the derivation. To take an example with "displacement" (in the transformational account), consider a sentence such as *She will think he was arrested* (Epstein's example 17). Here, *he* has by assumption been paired with the inflectional item set associated with *was*, and therefore c-commands all the terms of that item, namely, *arrested* and any NP objects of arrested.

Space prevents us from demonstrating how each of the fundamental grammatical relations displayed earlier can be similarly *derived* from the

[4]This important thesis was discovered first by Kitahara, 1994. We put to one side Bach's important points about "local" movements, which may well be phonological in nature, since they seem to involve adjacency.

[5]We note in passing, though, that the lack of precedence relations in the HC derivation lines also gives us "at no cost" the usual cases of quantifier ambiguity, without having to go through the usual (syntactic) contortions of segments, LF lowering, etc., as is common in so-called government-and-binding accounts: instead, we can simply leave the precedence relations *unstated* between two quantified NPs, such as *everyone* and *someone* in *everyone loves someone*—precisely because the syntax *must* leave them unstated. We may picture the NPs as "hanging off" a joint node ahead of the predicate, but this is to be considered as only a graphical aid.

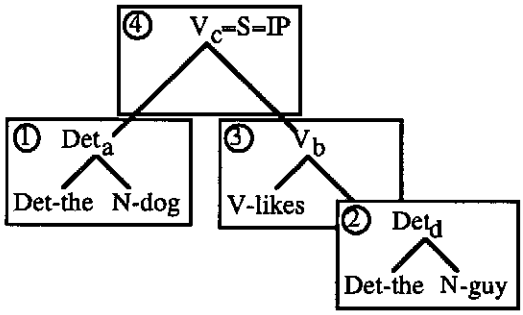
basic properties of HC, but the take-home moral is quite strong: (syntactic) grammatical relations are precisely those brought into existence by the syntactic compositional operator.

3 The Computational Implications: LR parsing

So far, we have talked about abstract derivations (proofs) of sentences from lexical multisets, interfacing to the motor/perceptual systems via some linearization process, yielding the left-to-right vocal (alternatively gestural) output that we hear. Turning this problem around, the hearer receives a temporally left-to-right ordered sound signal and, via its perceptual apparatus, recovers (at least conceptually), the hierarchical grammatical relations required in order to interface to the cognitive-meaning component of language—i.e., the view of language as (sound, meaning) pairs. To a first approximation then, the *parsing problem* becomes: how to recover from the *linearized* input that contains only precedence relations the *hierarchical* derivation lines.

Interestingly, a straightforward solution to this interface problem presents itself in a nearly "natural" formulation. Let us define a "natural" solution to the interface problem as follows: the perceptual/parsing system (conversely, the motor/articulatory or production) system should deliver input (alternatively output) to the linguistic faculty that can be easily read — that is, mapped to the same elements that the linguistic faculty employs, with minimal computational effort. In the best case, that mapping will simply be the identity function.

Now consider a canonical (i.e., natural) LR parse of a sentence such as *the dog likes the guy*. Recall that an LR parse constructs a rightmost derivation in reverse, working left-to-right. If one examines the order in which mergers or completions are built by an LR machine, it is easy to see that they mirror exactly the order of HC operations (in reverse). Moreover, canonical LR operation guarantees that the linguistic condition of strict cyclicity will be obeyed (that is, there can be no "interarboreal" operations that go back and modify to an item already shipped off to semantic interpretation — at least, not in the cases ordinarily considered. In this sense, the perceptual machinery meshes perfectly with the requirements of the language faculty — a surprising condition, if true.[6]



4 Conclusions from the Beer Hall

To summarize, given the current push towards "minimalism" in the so-called government-and-binding approach seems to have eliminated both government (and binding, not discussed here) in favor of a single hierarchical concatenation operator that meshes perfectly with the classical theory of categorial grammar, as well as providing a natural *explanation* for the observed grammatical relations and a transparent framework on which to build a model of sentence processing. While this trend surely does not solve all our "religious" problems, it certainly goes a long way towards taking down the "barriers" to a mature, ecumenical framework within which to reach common ground among what has long appeared to be quite disparate accounts of natural language. Perhaps we can all now drink beer together.

References

Abney, S. 1987. The English Noun Phrase in its sentential aspect. Ph.D. Thesis, MIT Department of Linguistics and Philosophy.

Bach, E., E., 1984. Some generalizations of categorial grammar. In F. Landman and F. Veltman (eds), Varieties of Formal Semantics. Dordrecht: Riedel, 1–23.

Bach, E., 1988. Categorial grammars as theories of language. In R. Oerhle, E. Bach, and

[6]Clearly, the language faculty and perceptual apparatus do *not* mesh in the familiar cases of sentences that are difficult-to-parse or produce — but this is a different sense than the "legibility" and "transparency" requirement intended in the main text.

D. Wheeler, Categorial Grammars and Natural Language Structures, Dordrecht: Kluwer, 17–34.

Chomsky, N., 1993. Bare phrase structure. In The View from Building 20. Cambridge, MA: MIT Press.

Chomsky, N., 1995. The Minimalist Program. Cambridge, MA: MIT Press.

Curry, H., 1961. Some logical aspects of grammatical structure. In R. Jakobson (ed)., Structure of Language and Its Mathematical Aspects, Proceedings of Symposia in Applied Mathematics, Providence, RI: American Mathematical Society, 56–68.

Epstein, S., 1995. Unprincipled syntax. Manuscript, June, 1995, Harvard University.

Kitahara, H., 1994. Target alpha: A Unified Theory of Structure Building. PhD. thesis, Harvard University Department of Linguistics.

McNaughton, J. and Papert, S., 1967. Counter-free Automata. Cambridge, MA: MIT Press.