Meeting the Boojum

Gregory M. Kobele
University of Chicago

There are a multitude of different linguistic theories on the market with sometimes quite different formal properties. This situation is beneficial to linguistics as a field for at least two reasons. First, different formal mechanisms make salient different sorts of generalizations, and the different communities have in addition different methodological commitments; this means that researchers from different traditions will tend to not only approach the same data differently, but will also see different patterns therein. In other words, having a wealth of communities with different modes of engaging with the data and with different prior beliefs, increases the chances of finding interesting generalizations.[1] Second, particularly given the mixed attitudes [5, 18] toward formal precision and mathematical rigor among linguists, the need to engage across notational boundaries forces some essential metatheoretical reflection about what exactly the commitments of one's analyses are; what is the theoretical wheat, and what the notational chaff? Müller and Wechsler engage precisely in this activity in the present paper. I think that this opens the door to a considered and rational discussion of the nature and content of the debate between 'phrasal and lexical' approaches to grammatical analyses. What I think is quite clear is that lack of formal precision is an impediment to this sort of discussion. Therefore, if one thinks (and I do) that such discussions should be had, then one should endeavour to make his or her theory more formally explicit; this is then a case where "inquiry is advanced by [. . .] fuller formalization" [5].

I view Müller and Wechsler's paper as having two facets. On the one hand, it is arguing against recent proposals of Tomasello and Goldberg, and is thus part of a continuing dialogue regarding the merits of two different styles of analysis (CxG *à la* Goldberg and HPSG *à la* Müller and Wechsler). The broader moral drawn however is that *lexical* approaches (to argument structure) are better than *phrasal* approaches. It is this broader claim that I will focus on in this response. There are three problems that I have in particular. First, it is not at all clear what the distinction Müller and Wechsler are trying to make is. Second, Müller and Wechsler's claims about simplicity favoring the lexical approaches are ungrounded. Finally, Müller and Wechsler's conclusions trade on an equivocation between 'analyses which make certain distinctions between expressions'

[1] I am making here an analogy to *hill-climbing* algorithms. One way to avoid the problem of getting stuck in a local maximum is to run the same algorithm multiple times, with different initial positions. I am suggesting that the different linguistic traditions play the role of different initializations of the 'algorithm' carried out by the scientist.

and 'HPSG *à la* Müller and Wechsler,' with the arguments establishing only that our analyses must make certain distinctions between expressions, which everyone already accepts.

1 On the lexical vs phrasal distinction

In trying to draw a conclusion broader than just "our analysis is better than Goldberg's," Müller and Wechsler recast this debate as one between "lexical approaches" and "phrasal approaches." But what exactly are these?

Intuitively, a (syntactic) analysis of two sentences is lexical just in case the difference between the sentences is present already at the word level; it is phrasal if there is no difference between the sentences at a word level, but there is one at the phrasal level. (It is sentential if the phrase level at which a difference is introduced is the sentence.) This is visualizable in terms of analysis trees. The difference between the active and passive sentences in figures one and two are present before the lexical item NIBBLE combines with either of its logical arguments.[2] A phrasal analysis of the passive [13] would introduce a difference only after the verb and its logical object had been combined. A sentential analysis [4] would introduce a difference after the verb had been combined with its subject as well.[3]

This almost makes the distinction that Müller and Wechsler are after; typical CG analyses are lexical in this sense. On the other hand, we see that there is no possible phrasal analysis of the passive in GPSG; once a constituent containing both the verb and logical object is created, the word order of the two is fixed.

So if this intuitive distinction is not what Müller and Wechsler intend, what could it be? In their abstract they state that

> In lexical approaches [...], lexical items [represent] essential information about potential argument selection and expression, but abstracts away from the actual local phrasal structure. In contrast, phrasal approaches [...] reject such lexical argument structures.

[2] Geometrically, there is no subtree containing both verb and object which is shared in figure one and two.

[3] Although intuitive, this distinction is problematic. Salvati [19] proves that Minimalist Grammars (MGs) [22] can be recast in the (meta-)grammatical framework of Abstract Categorial Grammar (ACGs) [7] using 'lowering' instead of movement. In other words, moved expressions are introduced into the containing expression at their highest moved-to positions. Thus, given a phrasal minimalist analysis of passive (i.e. one according to which both passive and active sentences are built from a constituent containing both verb and object), its ACG reformulation might very well not be phrasal in this sense. This is an awkward situation; the nature of the grammatical combinatory operations make writing phrasal analyses difficult, yet we can directly translate intuitively phrasal analyses into these terms. We see something similar with construction grammar, and its HPSG reformulation [2], as already noted by Müller and Wechsler (their footnote one). I think that the problem applying the intuitive diagnostic of lexicality vs phrasality to these other cases is that the intuitive diagnostic holds for tree-like structures, whereas Salvati's ACG encoding of a MG uses higher order abstract terms, and HPSG analyses make crucial use of reentrancy.

Everyone agrees that (descriptively speaking) different words require (and allow) different numbers of and types of dependents in different constructions. There is however some disagreement as to how best to account for this. By far the most common approach is to encode this information in terms of syntactic category, whether atomic (as in GPSG) or complex (as in CG). Another option, which underlies Borer's exoskeletal work, is to use a filter to exclude the undesired combinations. In many interesting cases, this filter can be expressed as a regular constraint (over derivation trees), and is just a notational variant of the syntactic category approach [9, 14].[4] As long as there are only finitely many categorial distinctions to be made, the atomic and complex syntactic category variants are formally equivalent. Müller and Wechsler may have in mind the claim that we cannot place an upper bound on the number of distinctions we need to make; they seem to argue in §8.1 that data from Turkish requires that we allow categories to represent an arbitrary number of desired dependents (but see Stabler [21] for a different interpretation). We can be a little more concrete if we assume that the syntactic structures in question are trees. Müller and Wechsler seem to be imagining a structure like the one on the left in figure 1, where the blue nodes labeled `LR` (for 'lexical rule') dominating the verb must be equinumerous with the red nodes labeled `A` (for 'argument') introducing arguments. Here we imagine that the nodes `LR` modify the form of the verb

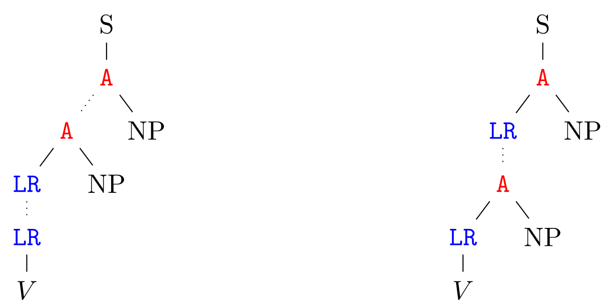

Figure 1: Different approaches to argument structure

in some way (perhaps by adding a causativizing affix). This family of trees is not *regular*, and thus cannot be recognized by a finite state tree automaton (see Comon et al. [6] for more information).[5] Thus if we want to assign structures like those on the left, we need to have a complex category system which allows us to remember however many `LR` nodes we have seen; this is what Müller and Wechsler's HPSG analysis looks like.

However, a rearrangement of the nodes in these trees, as per the tree on the right in figure 1, preserves the property of there being equal numbers of `LR`

[4] Another idea is to reduce this information to independently motivated semantic distinctions. This idea is extremely seductive, as many basic selectional properties are clearly related to semantic type (NPs denote generalized quantifiers, Ss propositions, etc), but no one has yet been able to make this work.

[5] It is a context-free tree language, and can be recognized by a push-down tree automaton.

and `A` nodes in each tree, but this family is *regular*, and can be recognized by a finite state tree automaton. Here, we have simplified the trees by turning the dependence between `LR` and `A` nodes into a *local* one. If we are willing to assign *these* structures instead, we can do just fine with a finite set of categories; this is what a typical minimalist analysis looks like.

So is the difference Müller and Wechsler are looking for one of atomic (phrasal) versus complex (lexical) categories? Or rather, since atomic and complex categories are notational variants as long as there are only finitely many of them, is the distinction finite versus infinite numbers of useful categories? Certainly, the GPSG and CG examples align with the atomic/phrasal vs complex/lexical distinction. On the other hand, CG only uses finitely many categories, and thus is a notational variant of a system with atomic categories.[6]

I do not see how to reconstruct a meaningful distinction between lexical and phrasal analyses which cuts the pie in the way Müller and Wechsler desire. Müller and Wechsler themselves note that it can be "difficult to distinguish" phrasal from lexical proposals, and that many are "basically similar to the lexical valence structures" they propose. In their paper, however, Müller and Wechsler actually only explicitly argue that HPSG-style lexical rules provide the right analysis of argument structure constructions. This might hold even if there should turn out to be no substantive lexical vs phrasal distinction.

2 Simpler than . . .

In §5.3, Müller and Wechsler argue that CxG is not, in fact, simpler than (HPSG with) lexical rules. They claim in particular that once CxG analyses explicitly record categorial and subcategorization information, they become as complex as the equivalent HPSG analyses. I think that this ignores an important aspect of complexity analysis.

We as linguists are not just writing analyses of particular languages, but also analyses of how languages vary.[7] The analyses of language variation can be given in terms of restrictions on the underlying grammar formalism ('universals'). Some grammar formalisms have restrictions built-in; of particular interest are the so-called 'mildly context-sensitive' grammar formalisms [11]. These include Tree Adjoining Grammars [12], Combinatory Categorial Grammars [23], and Minimalist Grammars [22]. These grammar formalisms only generate languages which are recognizable in polynomial time, and thus provide an account of the fact that only such patterns occur in human languages. In contrast, turing complete grammar formalisms like HPSG and Aspects-style Transformational Grammars generate all recursively enumerable languages [8, 17]. To the best of my knowledge, no substantive restriction on HPSG has been found which

[6] CCG [23], on the other hand, can use infinitely many categories, as does the Lambek Calculus [15].

[7] This latter may be reducible to the interplay of learning and language change, plus initial conditions.

admits popular HPSG analyses.[8] In addition, Carpenter [3] shows that lexical rules as used in HPSG are already by themselves sufficient to make the grammar formalism undecidable (equivalent to turing machines). Therefore, according to the most obvious meaningful comparison of grammar formalisms HPSG (with or without lexical rules) comes out as (vastly) more complex than these others.

A comparison of analyses written in different grammar formalisms is not complete unless the typological commitments of the formalisms are taken into account. Thus, even if it is true, as Müller and Wechsler suggest, that once the CxG analysis explicitly represents selectional restrictions, it becomes as complex as Müller and Wechsler's, it may still be less complex overall once the commitments of the underlying CxG framework (once made explicit) are taken into account.[9]

3 Lexical rules vs ...

Three challenges noted by Müller and Wechsler for analyses of argument structure revolve around the relation between morphology and syntax (the interactions between valence changing processes and derivational morphology and word coordination, and the iteration of valence changing processes). Müller and Wechsler note that, if your analysis involves first constructing the appropriate morphological forms of words, and then putting them in the appropriate order, you will need to somehow record the morphological construction process each word underwent, as this is relevant for determining their syntactic behaviour.[10] They observe that this information can be encoded in a lexical valence structure.

Another empirical challenge Müller and Wechsler pose is how to capture the distribution of arguments in partial fronting constructions. They note that if an analysis makes use only of string concatenation as a combinatory operation, then it must record which arguments of the verb appear next to it, and which have not, so as to enforce that all and only those which have not already appeared with the verb appear in the remainder of the clause. They observe that this information can be encoded in a lexical valence structure.

These observations both seem reasonable, but they (1) are dependent on assumptions (the lexical integrity principle[11] on the one hand, and that only

[8] Offline parsability constraints [10] make recognition decidable (though NP-hard), but rule out unary rules (which can apply to their own output); these are what Müller and Wechsler understand lexical rules to be.

[9] From a bayesian perspective, unrestrictive formalisms are allocating probability mass to grammars which will (*ex hypothesi*) never be needed.

[10] This is the situation as sketched in the tree on the left in figure 1.

[11] Ackema and Neeleman [1] describe the *Lexical Integrity Principle* (LIP) as stating that "*the internal structure of words is not accessible to the syntax.*" From an empirical perspective, the LIP is an attempt to explain a number of facts about language, such as that pieces of words tend not to be extractable, gappable, or conjoinable. (The explanation then being that this is because pieces of words are not subject to syntactic manipulation.) Rejecting the LIP is not a rejection of the data, but just a rejection of the particular explanation of this generalization that it offers. Much work in the transformational tradition (generative semantics, government & binding, antisymmetry, distributed morphology, nanosyntax) rejects the

concatenation of strings is available[12]) that are not universally accepted, and (2) establish conclusions (that we need to enforce distinctions between syntactic objects which encode the number and type of their dependents) which are uncontroversial.

3.1 Exoskeletalism

I claimed above that the conclusion of Müller and Wechsler's arguments, namely that we need to enforce distinctions between syntactic objects which encode the number and type of their dependents, was uncontroversial. Superficially, at least, recent approaches to syntax which view it as a 'free combinatorial system' (as in Borer's exoskeletal work), appear to be in conflict with this. However, as I alluded to in §1, this is a purely notational decision without any formal import. In grammar formalisms like MGs or TAGs, it is useful to distinguish between *derived* trees and *derivation* trees. (The derivation tree is a recipe for constructing the derived tree.) The reason for this distinction is that the derivation trees are formally simpler than the derived trees; indeed, the derivation trees form a regular set, while the derived trees do not. However, the derived trees can be obtained from the derivation trees by means of a tree transduction (a tree homomorphism with state). (The type of tree transduction depends on the grammar formalism in question.) Thus, the description of the set of observables (the (yields of the) derived trees) is simplified by factoring it into a description of a set of simpler but unobservable derivation trees, together with a description of a function mapping from derivation to derived trees [16]. (This perspective is generalized in ACGs [7].) There are two well-known results in the tree-automaton community which are relevant here. First, given a finite state tree transducer τ defined on trees over an input alphabet Σ, we can convert it to a tree homomorphism h (a single state tree transducer) over the richer alphabet $\Sigma \times Q$, where Q is the set of states of τ, so that $\tau(T_\Sigma) = h(R)$, where T_Σ is the set of all trees over alphabet Σ, and R is a regular subset of $T_{\Sigma\times Q}$. This fact, in linguistic terms, tells us that we can, by annotating our node labels with extra information, eliminate all computation from the transformation from derivations to derived trees, allowing this map to run blindly. In the other direction, we can, given a regular subset $R \subseteq T_\Sigma$, and a homomorphism h defined on R, encode the distinctions made in defining R into a finite set of states of a new tranducer

LIP, giving rise to a characteristic 'decompositional' style of analysis.

[12] Grammar formalisms which avail themselves of non-concatenative string operations make possible analyses in which local relations in the structures assigned to sentences do not correspond to local relations between substrings. Perhaps the most familiar sort of non-concatenative string operation to linguists is an operation which wraps a pair of strings $\langle u, v\rangle$ around another w to obtain uwv. A generalization of this treats the exponents of the grammar as tuples (of arbitrary finite length) of strings, and syntactic combination results in interleaving of tuples. Multiple Context-Free Grammars (MCFGs [20], also LCFRSs [24]) are a grammar formalism with desirable computational properties built around this kind of generalized wrapping operation. Diverse grammar formalisms such as Tree-Adjoining Grammar [12] and Minimalist Grammar [22] can be straightforwardly encoded as MCFGs, which means that the operations of these linguistic grammar formalisms (even movement) can be understood in terms of generalized string wrapping.

τ, such that $h(R) = \tau(T_\Sigma)$. Linguistically, this says that we can move the locus of the enforcement of syntactic dependencies from the category system itself to the 'interface maps.' In other words, we can freely get rid of syntactic categories by beefing up the interface maps. This holds as long as the derivation tree sets of your grammar formalism are regular, and the mapping from derivation trees to derived trees is a finite state transduction, as they are in MGs and TAGs (or, more generally, in second order ACGs).

4 Conclusions

Müller and Wechsler engage in the important undertaking of cross-framework dialogue. They are hampered in this regard by the lack of formal precision in the other theories and analyses they are considering. Other than encouraging more explicitness in linguistic analysis, their particular argument suffers from a number of difficulties. (1) It is not clear that the distinction they are arguing for is a useful or even meaningful one. (2) It is not clear that the metric of simplicity they implicitly appeal to in claiming that lexical rules *à la* HPSG are no more complex than alternatives is reasonable. (3) The arguments they marshal establish uncontroversial conclusions.

References

[1] Peter Ackema and Ad Neeleman. Syntactic atomicity. *Journal of Comparative Germanic Linguistics*, 6:93–128, 2002.

[2] Hans C. Boas and Ivan A. Sag, editors. *Sign-Based Construction Grammar*, volume 193 of *CSLI Lecture Notes*. CSLI Publications, 2012.

[3] Bob Carpenter. *The Logic of Typed Feature Structures*. Number 32 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, England, 1992.

[4] Noam Chomsky. *Syntactic Structures*. Mouton, The Hague, 1957.

[5] Noam Chomsky. On formalization and formal linguistics. *Natural Language and Linguistic Theory*, 8(1):143–147, 1990.

[6] Hubert Comon, Max Dauchet, Rémi Gilleron, Florent Jacquemard, Denis Lugiez, Sophie Tison, and Marc Tommasi. Tree automata techniques and applications. Available at `http://www.grappa.univ-lille3.fr/tata`, 2002.

[7] Philippe de Groote. Towards abstract categorial grammars. In *Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference*, pages 148–155, 2001.

[8] Nissim Francez and Shuly Wintner. *Unification grammars*. Cambridge University Press, New York, NY, 2012.

[9] Thomas Graf. Closure properties of minimalist derivation tree languages. In S. Pogodalla and J.-P. Prost, editors, *LACL 2011*, volume 6736 of *Lecture Notes in Artificial Intelligence*, pages 96–111, 2011.

[10] Efrat Jaeger, Nissim Francez, and Shuly Wintner. Unification grammars and off-line parsability. *Journal of Logic, Language and Information*, 14: 199–234, 2005.

[11] Aravind K. Joshi. Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In D. Dowty et al., editors, *Natural Language Processing: Theoretical, Computational and Psychological Perspectives*, pages 206–250. Cambridge University Press, NY, 1985.

[12] Aravind K. Joshi. An introduction to tree adjoining grammars. In A. Manaster-Ramer, editor, *Mathematics of Language*. John Benjamins, Amsterdam, 1987.

[13] Edward Keenan. Passive is phrasal not (sentential or lexical). In T. Hoekstra et al., editors, *Lexical Grammar*, volume 3 of *Publications in Language Sciences*, chapter 7, pages 181–214. Foris Publications, Dordrecht, 1980.

[14] Gregory M. Kobele. Minimalist tree languages are closed under intersection with recognizable tree languages. In S. Pogodalla and J.-P. Prost, editors, *LACL 2011*, volume 6736 of *Lecture Notes in Artificial Intelligence*, pages 129–144, 2011.

[15] Joachim Lambek. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170, 1958.

[16] Frank Morawietz. *Two-Step Approaches to Natural Language Formalisms*, volume 64 of *Studies in Generative Grammar*. Mouton de Gruyter, 2003.

[17] P. Stanley Peters and Robert W. Ritchie. On the generative power of transformational grammar. *Information Sciences*, 6:49–83, 1973.

[18] Geoffrey K. Pullum. Formal linguistics meets the Boojum. *Natural Language and Linguistic Theory*, 7(1):137–143, 1989.

[19] Sylvain Salvati. Minimalist grammars in the light of logic. In S. Pogodalla et al., editors, *Logic and Grammar: Essays Dedicated to Alain Lecomte on the Occasion of his 60th Birthday*, volume 6700 of *Lecture Notes in Artificial Intelligence*, pages 81–117. Springer, 2011.

[20] Hiroyuki Seki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. On multiple context-free grammars. *Theoretical Computer Science*, 88:191–229, 1991.

[21] Edward P. Stabler. The finite connectivity of linguistic structure. In C. Clifton et al., editors, *Perspectives on Sentence Processing*, pages 303–336. Lawrence Erlbaum, NJ, 1994.

[22] Edward P. Stabler. Derivational minimalism. In C. Retoré, editor, *Logical Aspects of Computational Linguistics*, volume 1328 of *Lecture Notes in Computer Science*, pages 68–95. Springer-Verlag, Berlin, 1997.

[23] Mark Steedman. *The Syntactic Process*. MIT Press, 2000.

[24] K. Vijay-Shanker, David Weir, and Aravind Joshi. Characterizing structural descriptions produced by various grammatical formalisms. In *Proceedings of the 25th Meeting of the Association for Computational Linguistics*, pages 104–111, 1987.