The epicenter of linguistic behavior

Edward P. Stabler

Abstract

Speaking, listening, reading, and reasoning all depend on common, fundamental aspects of linguistic structure. This common structure is defined by the grammar, and so Bever (1970) calls grammar 'the epicenter of all language behavior' in his pioneering study of how grammar and the structures it defines may relate to other cognitive abilities. Computational models with this perspective have become more sophisticated since 1970. While many fundamental questions remain open, a substantial, non-trivial consensus about the nature of linguistic structure has emerged, and a wide range of methods for computing that structure are well understood. Much less is established about how the structure or content of grammar may be grounded in other cognitive abilities, but here again there is a much better understanding of how these *could* be related, and some proposals are getting sophisticated enough for preliminary empirical assessments.

The influence of language on normal human cognition is vast and varied. We hear, speak, read, write, notice clever turns of phrase, chant incantations, sing and dance to rhymes. The idea that all this activity is governed by just one, uniform cognitive mechanism is plainly a non-starter, and yet in all these activities the distinctively linguistic structure is largely shared. For Bever (1970), this common structure is the 'epicenter of language', the grammar, and the project of identifying its role in mental life has been a fundamental part of constructing "an experimental mentalism" (Fodor, Bever, and Garrett, 1974, p.xi). But in the past decades, besides all the expected vicissitudes of experimental science, this project has faced a number of conceptual obstacles. Some of the most significant of these are finally tumbling down. In particular, there are at least three conceptual issues which have been yielding as computational studies of language have advanced in the past couple of decades. *First*, languages are vast and enormously various. Getting a grip on their role in mental life depends on understanding something about the properties they all have. Attempts to obtain a broader view by listing their significant, universal properties, so prominent in the 1960's and 1970's, seem to many to have foundered. Languages that are consistently OV tend to be agglutinating (Lehmann, 1973), but there is the fusional Akkadian; languages in which the demonstrative follows the noun tend to have the adjective after the noun too (Hawkins, 1983; Greenberg, 1963), but there are Hausa, Tzotzil and many other apparent exceptions. Literally thousands of such proposals and apparent counterexamples are now documented.[1] Chomsky (1965, p.209n; 1976, p.56) suggests that such proposals are not looking for properties that are 'deep' or 'abstract' enough, considering for example the restrictions on extractions from subject positions, and noting that the bearing of evidence on such abstract properties can only be assessed relative to a "pretty rich framework of analytic assumptions." But this raises more conceptual problems!

Turning to the rich framework of theoretical assumptions needed to relate relatively abstract principles of grammar to linguistic judgments about particular sentences or any of the other various linguistic behaviors, we face a *second* fundamental issue: it seems there is no consensus on even the most basic theoretical claims. Are sentence structures built in part by movements that leave traces? And if there is movement, do we really have head movement, A-movement, covert movement, and a special parallel merge for coordinate structures? Are structures built up from lexical items whose pronunciation is then altered by phonological rules, or does an independent level of morphological structure intervene with its own well-formedness conditions? Or are there no rules at all, but only preferences determining an optimal form holistically? Even within a particular linguistic tradition that may have some loose consensus, the points of consensus seem to shift constantly. As Townsend and Bever (2001, p.45) note at the very beginning of a review of this literature, "especially in the last two decades, the rapid changes in syntactic theories have left psychologists in large part baffled as to how to integrate grammatical knowledge and behavior in rigorous models."

A *third*, related problem concerns the identification of the psychological mechanisms involved in computing grammatical structures. Even if we knew what the structures were, exactly, there seem to be a great range of possible ways to compute them. To put the matter more pointedly, even given a particular, grammatically-defined relation between pronounced forms and structures, it seems that there are so many degrees of freedom in computing any such relation that any possible evidence about judgments of well-formedness, self-paced reading times, fMRI results, etc. could be explained by a diverse range of computational models.

These three problems – what are the common properties of human languages; what assumptions appropriately relate grammar and various aspects of performance; and how could particular kinds of computations of those relations be evidenced? – these problems have encumbered the first steps toward a science of language and cognition in ways that are, I think, unprecedented in the sciences. The first problem is unique in the nature of the diversity to be tamed. The latter problems are unique to computational models of complex naturally occurring systems like the human mind.

1 Syntactic structure: revealing the hidden consensus

The sometimes rancorous debates at the frontiers of an active science can hide the common ground on which almost everyone stands. Of course, science does not require complete agreement about anything, but the bearing of evidence should be plain enough, even across competing scientific traditions, to have some force. In linguistic theory, a long series of results has revealed that the points of consensus are very much more substantial than linguists have realized. In particular, a very significant computational consensus was identified by Joshi (1985) in his hypothesis that human languages are both strongly and weakly mildly context sensitive (MCS). While any empirical test of this hypothesis still depends on a network of theoretical assumptions, the claim is so fundamental that it can be connected to many diverse traditions in grammar. To say that language is "strongly and weakly" MCS is to say that MCS grammars can both define the sentences of human languages (weak adequacy) and also provide the structures of those languages (strong adequacy). Joshi's original definition of MCS grammars was partly informal, so there are now various precise versions of his claim. One is that human languages are defined by tree adjoining grammars (TAGs) or closely related grammars, and another theoretically weaker (and hence empirically stronger) position is that human language are definable by the more expressive (set local) multi-component TAGs or closely related grammars. The most remarkable thing about this claim came out of the innocent-sounding phrase "or closely related grammars," because it was discovered that a wide range of independently proposed grammar formalisms falls under that description. In particular, a series of papers beginning in the 1980's and 1990's established the following inclusion relations among the languages defined by various kinds of grammars, across traditions:[2]

$$\text{CFG} \subset \boxed{\text{TAG}=\text{CCG} \subset \text{MCTAG}=\text{ACG}_{2,4}=\text{MCFG}=\text{MG}} \subset \text{CSG},$$

where the acronyms represent languages definable by: context free grammar (CFG); tree adjoining grammar (TAG); a certain kind of combinatory categorial grammar (CCG); set-local multi-component (MCTAG); a certain kind of abstract categorial grammar ($\text{ACG}_{2,4}$); multiple context free grammar (MCFG) – a restricted form of Pollard's (1984) generalized phrase structure grammars; minimalist grammar (MG); and context sensitive grammar

(CSG). A more recent series of papers has established a convergence internal to the Chomskian tradition in syntax:[3]

MG=MGH=DMG=CMG=PMG=SMMG=RMG,

where the acronyms represent languages definable by: minimalist grammar (MG); MG with head movement (MGH); MG with head parameters specifying whether complements are to the left or right (DMG); MG in which the features triggering (external) merge can also trigger move (internal merge) (CMG); MG in which phases constrain extraction (PMG); MG with sideward movement (SMMG); MG with in which relativized minimality constrains extraction (RMG). Every linguist knows that these grammars were not designed to mimic each other. On the contrary, these various ideas come from sometimes fiercely opposed grammatical traditions, and yet they converged on grammars that are weakly equivalent in the sense that they define exactly the same sets of sentences. Not only that, but many of these proofs of weak equivalence are easy because the recursive mechanisms are actually much more similar than superficial differences would suggest. To take one dramatic example, MGs were designed to formalize fundamental ideas from the Chomskian tradition in syntax, but for every MG there is a weakly equivalent MCFG which is also strongly equivalent in the sense that there is an isomorphism between the derivation trees for every sentence generated by the grammars.[4]

It takes some work to really appreciate the significance of these results. Probably the best known fact about the MCS languages is that while they seem expressive enough to define the discontinuous dependencies found in human languages, with some few exceptions under investigation,[5] they can be recognized efficiently in the computer scientists' sense; that is, a polynomial function sets an upper limit on the number of steps required to correctly decide whether a sentence is generated by the grammar. This property may not have been expected, since garden paths and various other kinds of constructions are not successfully recognized by people even in short sentences. A *second*, more important but much less well-known property of many MCS grammars is that, while they may derive rather complex structures, the derivations themselves are relatively simple. This was noticed early for TAGs (Vijayashanker, 1987; Weir, 1988) and more recently for MGs (Michaelis, Mönnich, and Morawietz, 2000; Morawietz, 2001; Kobele, Retoré, and Salvati, 2007; Mönnich, 2007; Graf, 2011; Kobele, 2011). These grammars can be regarded as having similar 'two step' structures: a relatively simple derivation feeds an even simpler mapping to pronounced, interpreted forms. Consider the simple tree on the left in Figure 1, which is built by the derivation depicted on the right. The structure of the derivation is apparently simpler than

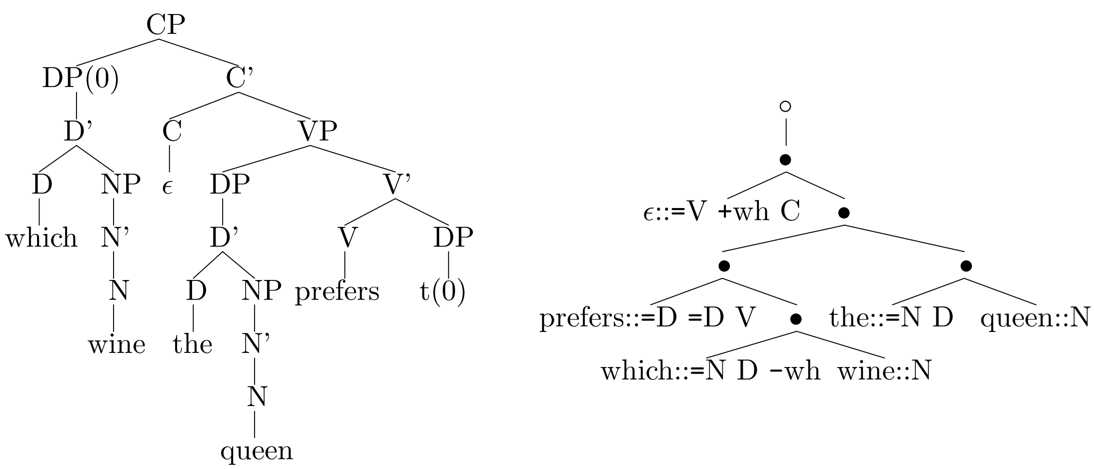

Figure 1: A derived tree and its MG derivation. Here, • represents an (external) merge step, and ○ represents a movement (an internal merge). See Stabler (2012) for the parsing algorithm that computed these structures from the input *which wine the queen prefers.*

the tree it derives, and when this idea is pinned down the difference is clear and dramatic. Unlike the derived trees, the derivation trees on the right allow a certain kind of finite state definition; the derivation trees form a regular tree set. The mapping from derivations to derived trees can be done by a single deterministic traversal (by a 'multiple bottom-up tree transducer'), and we gain an important insight into how the pronounced languages could be parsed: to find the derivation, we must in effect undo the deformation provided by the mapping to surface order. But that mapping is extremely simple in *all* MGs and also in the many MG variants surveyed by Stabler (2010), for example.

Identifying the basic structural properties of MCS languages allows us to dismiss superficial differences in the notations and formalisms used to define them. Consider for example the question of whether there are actually traces in syntactic structure. Movement operations are something special, but recognizing them does not require putting traces into the structure. In fact, this idea is an inessential, stipulated detail which can be regarded one way or another without significant consequence for the most basic assumptions about language structure. So the current diversity of opinions about them is no surprise. Some linguists like traces, others prefer 'multidominance structures', and others regard the representation of derived structure at the interfaces as the special consequences of a fundamentally very simple derivation structure. These conceptions may turn out, after all, to differ in ways that matter, but the derivational operations themselves do not depend on them.

A *final* development coming out of the MCS convergence is new attention to learning methods for substantial subsets of these languages (Yoshinaka and Clark, 2010; Yoshinaka, 2010). This preliminary and more recent ongoing work is very promising.

All these computational results are notably different in character from the speculations on the front line of linguistic research: these results pertain to very large classes of grammars, grammars with properties that most reasonable linguists would accept or at least recognize as very similar to what they do accept. This consensus is stable and rather well understood.

2 Performance models: basic properties

Considering how a certain sort of 'analysis-by-synthesis' parsing method might fit with recent work in the minimalist program of Chomskian syntax, Townsend and Bever (2001, p.179) say, "Hope springs eternal: perhaps a new 'derivational theory' of the psychological operations involved in assigning derivations is at hand." The formal studies mentioned just above have, at least, provided many ways that any of the MCS languages can be analyzed, and many of them are quite simple and direct. The first and perhaps easiest methods discovered were all-paths 'chart parsing' strategies, now well reviewed in a text by Kallmeyer (2010). These methods have not been so appealing to psychologists, though, since (i) they require more and more memory, without bound, to analyze longer sentences, and (ii) they do not single out particular structures in a way that would predict garden path effects. These two properties may favor automata-theoretic and one-path-at-a-time parsing methods (which can be provided with reanalysis possibilities using a 'backtrack stack' or 'beam' of alternative analyses). These too have been developed, with some particularly transparent and simple proposals appearing recently (Stabler, 2011; Stabler, 2012),[6] following on the structural insights mentioned earlier. These parsers can recognize arbitrarily long sentences of certain simple kinds with just finite memory, and they develop a favored analysis at each point. They are extremely simple, but easily recover correctly even the rather elaborate derivations of remnant-movement analyses. These (or the similar analyses of any of the MCS alternatives) should be on the list of what every psychologist should know about how linguistic structure could be calculated. The analysis-by-synthesis approaches favored by Bever and others (Bever and Poeppel, 2010) are very similar to some of the recently emerging 'rational' computational approaches (Hale, 2011) which are now being extended to MCS languages – these are not necessarily top-down, but capable of being nearly so in the recognition of clear, fluent speech.

3 Habits of phrasing

One thing that has been clear since the earliest probabilistic studies of syntax (Miller and Chomsky, 1963; Grenander, 1967) is that the factors determining which expressions are most likely in a given context are widely various and poorly understood compared to the much narrower determinants of grammatical properties. Language architectures which do not recognize the distinction face the difficult problem of explaining the obvious facts: what you just said has an enormous influence on what I am likely to say in reply, in ways that cross-cut the narrow and local constraints of grammar. The universal bounds imposed by grammar may go significantly beyond Joshi's mild context sensitivity – opinions significantly enriching that universal are mainly controversial and shifting, in ways that are completely unsurprising. But the MCS grammars define units of structures over which preferences, for various diverse linguistic tasks, are naturally defined (Hale, 2006; Joshi and Sarkar, 2002).

4 Computational perspectives on the epicenter

Computational methods provide tools for describing rather abstract similarities of structures and languages. Most discussions of language universals stick entirely to 'concrete universals' to the extent that they do not even notice the very strong and remarkably uncontroversial computational properties of all the serious contenders in linguistic theory (or at least, those contenders that give any attention to the project of providing adequate descriptions in a restricted formalism). The first stumbling block of inappropriate superficiality is removed with this realization. Computer scientists are familiar with the fact that there are many ways to implement a computation; and mathematicians know that there are many ways to define a set of structures. Linguists and psychologists must also accommodate these facts, to recognize that in spite of the many ways to define a language, there are also many ways that will not work. It is important to know the difference. A second stumbling block of getting lost in the shifting diversity of theoretical assumptions is significantly reduced by attention to what the significant differences are. One rather surprising recent development in the study of mildly context sensitive grammars has been the emerging understanding of their 'two-step' character: derivation + mapping to derived and pronounced forms. The details of derived structures are much less important than the derivations themselves and their connections to pronounced forms. The simpler structure we find when derivations are isolated catalyzes the study of how simple the mechanisms of analysis might really be, and of how those mechanisms could extend to, or even across, interfaces. The third stumbling block of knowing how to look for reasonable implementations is reduced by comparing alternatives that really differ significantly: e.g. those that require unbounded memory vs. those that do not; those that predict left-branching to be as difficult as center-embedding; etc. The relevant fault lines that separate feasible, empirically defensible models from the rest become clear as the broader landscape of alternatives comes into view.

References

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In J.R. Hayes, editor, Cognition and the Development of Language. Wiley, NY.

Bever, Thomas G. and David Poeppel. 2010. Analysis by synthesis: A re-emerging program of research for language and vision. Biolinguistics, 4(2).

Bhatt, Rajesh and Aravind Joshi. 2004. Semilinearity is a syntactic invariant: a reply to Michaelis and Kracht. Linguistic Inquiry, 35:683–692.

Chesi, Cristiano. 2007. An introduction to phase-based minimalist grammars: Why move is top-down from left-to-right. Technical report, Centro Interdepartmentale di Studi Cognitivi sul Linguaggio.

Chomsky, Noam. 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, Massachusetts.

Chomsky, Noam. 1976. On the nature of language. In S.R. Harnad, D. Steklis, and J. Lancaster, editors, Origins and Evolution of Language and Speech, Annals of the New York Academy of Sciences, Volume 280. Annals of the New York Academy of Sciences, NY, pages 46–57.

Fodor, Jerry A., Thomas G. Bever, and Merrill F. Garrett. 1974. The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. McGraw-Hill, NY.

Fong, Sandiway. 2005. Computation with probes and goals: A parsing perspective. In Anna Maria Di Sciullo, editor, UG And External Systems: Language, Brain And Computation. John Benjamins, Philadelphia.

Gärtner, Hans-Martin and Jens Michaelis. 2007. Some remarks on locality conditions and minimalist grammars. In Uli Sauerland and Hans-Martin Gärtner, editors, Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics. Mouton de Gruyter, NY, pages 161–196.

Graf, Thomas. 2011. Closure properties of minimalist derivation tree languages. In Logical Aspects of Computational Linguistics, LACL'11.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg, editor, Universals of Language: report of a conference held at Dobbs Ferry, New York, April 13-15, 1961. MIT Press, Cambridge, Massachusetts.

Grenander, Ulf. 1967. Syntax-controlled probabilities. Technical report, Brown University, Providence, Rhode Island.

Hale, John T. 2006. Uncertainty about the rest of the sentence. Cognitive Science, 30(1):609–642.

Hale, John T. 2011. What a rational parser would do. Cognitive Science, 35(3):399–443.

Harkema, Henk. 2001. A characterization of minimalist languages. In Philippe de Groote, Glyn Morrill, and Christian Retoré, editors, Logical Aspects of Computational Linguistics, Lecture Notes in Artificial Intelligence, No. 2099, pages 193–211, NY. Springer.

Hawkins, John A. 1983. Word Order Universals. Academic Press, NY.

Joshi, Aravind K. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In D. Dowty, L. Karttunen, and A. Zwicky, editors, Natural Language Processing: Theoretical, Computational and Psychological Perspectives. Cambridge University Press, NY, pages 206–250.

Joshi, Aravind K. and Anoop Sarkar. 2002. Tree-adjoining grammars and their application to statistical parsing. In R. Bod, R. Scha, and K. Sima'an, editors, Data Oriented Parsing. CSLI Publications, Stanford, pages 255–283.

Kallmeyer, Laura. 2010. Parsing beyond context-free grammars. Springer, NY.

Kanazawa, Makoto and Sylvain Salvati. 2007. Generating control languages with abstract categorial grammars. In L. Kallmeyer, P. Monachesi, G. Penn, and G. Satta, editors, Proceedings of the 12th Conference on Formal Grammar (FG'07), Stanford, California. CLSI Publications.

Kobele, Gregory M. 2006. Generating Copies: An Investigation into Structural Identity in Language and Grammar. Ph.D. thesis, UCLA.

Kobele, Gregory M. 2011. Minimalist tree languages are closed under intersection with recognizable tree languages. In Logical Aspects of Computational Linguistics, LACL'11.

Kobele, Gregory M., Christian Retoré, and Sylvain Salvati. 2007. An automata-theoretic approach to minimalism. In J. Rogers and S. Kepser, editors, Model Theoretic Syntax at 10, ESSLLI'07.

Lehmann, Christian. 1973. Worstellung in Fragesätzen. In Hansjakob Seiler, editor, Linguistic workshop I. Vorarbeiten zu einem Universalienprojekt. Fink, Munich, pages 20–53.

Michaelis, Jens. 1998. Derivational minimalism is mildly context-sensitive. In Proceedings, Logical Aspects of Computational Linguistics, LACL'98, pages 179–198, NY. Springer.

Michaelis, Jens. 2001a. On Formal Properties of Minimalist Grammars. Ph.D. thesis, Universität Potsdam. *Linguistics in Potsdam 13*, Universitätsbibliothek, Potsdam, Germany.

Michaelis, Jens. 2001b. Transforming linear context free rewriting systems into minimalist grammars. In Philippe de Groote, Glyn Morrill, and Christian Retoré, editors, Logical Aspects of Computational Linguistics, Lecture Notes in Artificial Intelligence, No. 2099, pages 228–244, NY. Springer.

Michaelis, Jens. 2002. Notes on the complexity of complex heads in a minimalist grammar. In Proceedings of the 6th International Workshop on Tree Adjoining Grammars and Related Frameworks, TAG+6, pages 57–65.

Michaelis, Jens. 2004. Observations on strict derivational minimalism. Electronic Notes in Theoretical Computer Science, 53:192–209. Proceedings of the Joint Meeting of the 6th Conference on Formal Grammar and the 7th Meeting on Mathematics of Language (FG-MOL '01), Helsinki, 2001. Available http://www.sciencedirect.com/science/journal/15710661.

Michaelis, Jens and Marcus Kracht. 1997. Semilinearity as a syntactic invariant. In Christian Retoré, editor, Logical Aspects of Computational Linguistics, pages 37–40, NY. Springer-Verlag (Lecture Notes in Computer Science 1328).

Michaelis, Jens, Uwe Mönnich, and Frank Morawietz. 2000. Derivational minimalism in two regular and logical steps. In Proceedings of the 5th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+5), pages 163–170.

Miller, George A. and Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush, and Eugene Galanter, editors, Handbook of Mathematical Psychology, Volume II. Wiley, NY, pages 419–492.

Mönnich, Uwe. 2007. Minimalist syntax, multiple regular tree grammars, and direction preserving tree transductions. In James Rogers and Stephan Kepser, editors, Model Theoretic Syntax at 10. ESSLLI'07 Workshop Proceedings.

Morawietz, Frank. 2001. Two Step Approaches to Natural Language Formalisms. Ph.D. thesis, University of Tübingen.

Pollard, Carl. 1984. Generalized phrase structure grammars, head grammars and natural language. Ph.D. thesis, Stanford University.

Radzinski, Daniel. 1991. Chinese number names, tree adjoining languages and mild context sensitivity. Computational Linguistics, 17:277–300.

Rambow, Owen. 1994. Formal and Computational Aspects of Natural Language Syntax. Ph.D. thesis, University of Pennsylvania. Computer and Information Science Technical report MS-CIS-94-52 (LINC LAB 278).

Salvati, Sylvain. 2011. Minimalist grammars in the light of logic. In Sylvain Pogodalla, Myriam Quatrini, and Christian Retoré, editors, Logic and Grammar, number 6700 in Lecture Notes in Computer Science. Springer.

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

Stabler, Edward P. 1997. Derivational minimalism. In Christian Retoré, editor, Logical Aspects of Computational Linguistics. Springer-Verlag (Lecture Notes in Computer Science 1328), NY, pages 68–95.

Stabler, Edward P. 2001. Recognizing head movement. In Philippe de Groote, Glyn Morrill, and Christian Retoré, editors, Logical Aspects of Computational Linguistics, Lecture Notes in Artificial Intelligence, No. 2099. Springer, NY, pages 254–260.

Stabler, Edward P. 2003. Comparing 3 perspectives on head movement. In A. Mahajan, editor, From Head Movement and Syntactic Theory, UCLA/Potsdam Working Papers in Linguistics. UCLA, pages 178–198.

Stabler, Edward P. 2010. Computational perspectives on minimalism. In Cedric Boeckx, editor, Oxford Handbook of Minimalism. Oxford University Press, Oxford, pages 617–641.

Stabler, Edward P. 2011. Top-down recognizers for MCFGs and MGs. In F. Keller and D. Reitter, editors, Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL), 49th Annual Meeting of the Association for Computational Linguistics.

Stabler, Edward P. 2012. Bayesian, minimalist, incremental syntactic analysis. Topics in Cognitive Science, forthcoming.

Townsend, David J. and Thomas G. Bever. 2001. Sentence Comprehension: The Integration of Rules and Habits. MIT Press, Cambridge, Massachusetts.

Vijay-Shanker, K. and David Weir. 1994. The equivalence of four extensions of context free grammar formalisms. Mathematical Systems Theory, 27:511–545.

Vijayashanker, K. 1987. A Study of Tree Adjoining Languages. Ph.D. thesis, University of Pennsylvania.

Weir, David. 1988. Characterizing Mildly Context-Sensitive Grammar Formalisms. Ph.D. thesis, University of Pennsylvania, Philadelphia.

Yoshinaka, Ryo. 2010. Polynomial time learning of some multiple context-free languages with a minimally adequate teacher. In Proceedings of the 10th International Colloquium on Grammatical Inference.

Yoshinaka, Ryo and Alexander Clark. 2010. Polynomial time learning of some multiple context-free languages with a minimally adequate teacher. In Proceedings of the 15th Conference on Formal Grammar.

Notes

[1]See "The Universals Archive" http://typo.uni-konstanz.de/archive/.

[2]Vijay-Shanker and Weir (1994), Seki et al. (1991), Stabler (1997), Michaelis (1998; Michaelis (2001b), Harkema (2001), Salvati (2011).

[3]Michaelis (2001a; Michaelis (2004; Michaelis (2002), Kanazawa and Salvati (2007), Gärtner and Michaelis (2007), Stabler (2001; Stabler (2003; Stabler (2010).

[4]Although every MG has a strongly equivalent MCFG, Stabler (2012) shows why the difference between them is significant. MGs explicitly mark distinctions that are neglected in the strongly equivalent MCFGs, and consequently can provide a starting point for feasible performance models in a way that MCFGs cannot.

[5]Radzinski (1991), Rambow (1994), Michaelis and Kracht (1997), Bhatt and Joshi (2004), Kobele (2006).

[6]Cf. also the more speculative Fong (2005), Chesi (2007).