Meta-meta-linguistics

Edward P. Stabler

Stokhof and van Lambalgen [40] suggest that modern linguistics has failed to meet the expectations one might have for it, that no consensus about central issues has emerged, and that this raises the fundamental question of whether modern linguistics is a failed discipline. They diagnose the problem as arising from treating language as grammar and linguistic competence as individual ability. Unlike the abstraction in the natural sciences, they say, these idealizations are not rational but ideological. The Stokhof and van Lambalgen argument is unpersuasive, though, because their expectations for the field are rather peculiar, their search for consensus is inappropriately superficial, and their diagnosis compares linguistic idealizations to an unrealistic conception of how abstraction works in the sciences.

This commentary aims to briefly remind the reader of a more optimistic, traditional perspective, one that is proposed in the literature cited by Stokhof and van Lambalgen but whose main points are oddly neglected by them. I think this traditional conception fits the facts much better.[1] §1 begins by accepting the idea advanced by Stokhof and van Lambalgen, namely that modern linguistics relies on idealizations that differ in fundamental respects from a simple conception of how abstraction works in physics. But this has been noticed many times before. That is, linguistic idealization does not work like the simple abstraction described by Stokhof and van Lambalgen, but neither does idealization in any of the sciences. The kind of idealization we find in linguistics is not "primarily ideological," but a rational strategy shared by the biological sciences, and perhaps even by all non-fundamental sciences, including most of physics. The empirical motivation for this strategy in linguistics is clear, and depends only on rather weak and plausible assumptions. The most likely failures of these assumptions are not the ones mentioned by Stokhof and van Lambalgen, and would not present the same symptoms. §2 agrees that we do not understand much about how linguistic structure is computed in any particular kind of linguistic "performance," or how language processing realized in the brain, neurophysiologically. But these are models that attempt to define the interaction of many different sorts of factors, and models of that kind are difficult in all the sciences. I think that we have seen very significant recent advances in performance models, as I will briefly mention, but the ultimate goals remain distant, as do the comparable goals in other cognitive domains. §3 observes some significant instances of consensus in linguistic theory, including some instances which the apparent diversity of linguistic traditions initially concealed.

1 Abstraction in the special sciences

The heart of the Stokhof and van Lambalgen [40] proposal is a critique of the way abstraction or idealization works in modern linguistics; they think it is "primarily ideological" and that it insulates linguistic theory from any relevant evidence.[2] This is a mistake.

[1] The term 'modern linguistics' is not pinned down by Stokhof and van Lambalgen [40]. Certainly they intend it to encompass a very large range of views in generative syntax (and generative phonology?), perhaps with the many variants of Chomskian syntax at its core. I will try to use the term in the same sense.

[2] I will not try to carefully follow Stokhof and van Lambalgen's special uses of the terms "abstraction" and "idealization," but will use them interchangeably.

Stokhof and van Lambalgen suggest that, in physics, when we consider frictionless planes or perfect vacuums, for example, "the phenomenon from which we abstract is a real one, and its reality is acknowledged in the theory or in the model that is based on the abstraction"; "a quantitative parameter of a phenomenon that is subject to abstraction is assigned a specific value (zero, infinite,...)." Linguistic competence, on the other hand, strips away "memory limitations, mistakes, (communicative) goals, attention shifts, and so on... In that way a new object of study is created, i.e. an object that has an ontological status different from the original one." Citing *Aspects* [9] and other works by Chomsky, Stokhof and van Lambalgen suggest that no rational ground for the latter sort of idealization has been offered, and that it is primarily an ideological move. Furthermore, they suggest that it differs fundamentally from abstraction in physics in that "the model itself does not contain any suggestion or clue as to how it could be related in the end to what we can in fact observe. In that sense, idealizations don't make life any easier, on the contrary, they create a lot of extra work."

This analysis of idealization in linguistics is very odd, since, in the first place, the rational basis for it is suggested by Chomsky on the very pages of *Aspects* they cite, and also in many other works. And in the second place, even if neither Chomsky nor anyone else had ever mentioned any rational basis for distinguishing competence in this way, a moment's reflection would suffice to suggest one. The linguistic idealization is apparently grounded in the empirical assumption that the mechanisms responsible for determining how phrases are formed in human languages are relatively independent of those involved in determining memory limitations, mistakes, attention shifts, and so on. So, as Chomsky and others put it, the proposal is that linguistic performance can be factored into these various interacting mechanisms. This is an instance of the common methodology that Plato (Phaedrus 265d-266a) describes as the attempt to carve up the natural world "at its joints." That is, we aim to find domains with causal interactions that are relatively closed, domains that can be described relatively autonomously. We see this method clearly in the biologists' attempts to describe the structure and function of organs of the body, and of species in the population of organisms. In psychology, we see it in attempts to separately describe the visual system, the auditory system, various systems of motor control, etc. And contrary to the suggestion of Stokhof and van Lambalgen, this same strategy is common in physics.

Stokhof and van Lambalgen suggest that, in physics and the other natural sciences, when we factor a phenomenon into various interacting components, we do so in a way that makes it clear how the factored parts interact. In fact, they suggest that the parameters factored away are actually given in the quantitative models of natural science, set to some limiting value (0 or infinity or whatever). Is that true? It is plain to see that this is a mistake. In grade school physics, we are all taught a simple formula for computing the period of a frictionless two-dimensional point-mass pendulum, on a rigid massless rod, under the influence of a constant gravitational force. In this setting we can get some version of the Galilean principle that the frequency of the pendulum is independent of its amplitude. But it is a safe bet that your grade school teacher was not prepared to provide a model of the pendulum when the idealizing conditions fail, and neither was Galileo. In university physics, we meet the nonlinear pendulum, much more challenging and still very highly idealized. Are we ever given a model in which we can see that our idealizations simply set some of physical parameters to limiting values? No, not even close! So, can it be done? In fact, it is not at all clear how to do this, and philosophers of science have been worrying about exactly this issue for quite a few years: even in very simple cases we do not know how to do this.[3]

So linguists are by no means alone in attempting to factor natural phenomena into interacting domains, typically without clear ideas about how all the factors interact. Nevertheless, we can raise directly the question of whether the motivation for the concept of linguistic competence is well founded. That is: could

[3]See, for example, [8], [24], [10] and other work on "ceteris paribus" conditions in physics and other sciences.

we be wrong in thinking that that the mechanisms responsible for determining how phrases are formed in human languages are likely to be relatively independent of those involved in determining memory limitations, mistakes, attention shifts, and so on? Of course, it is possible that we are wrong about this. What would happen in that case? We would expect to be unable to formulate stable generalizations about how phrases are built, for example, without reference to memory, attention and so on. Fortunately, syntax does not seem to have that problem.

The Stokhof and van Lambalgen worry about linguistic methodology seems to derive from a more general concern that linguistic theory is not responsive to evidence, because we do not know how grammar relates to performance. This worry could have been raised directly, and has been raised directly in much other literature. Obviously, linguistic theory is not based directly on performance models; rather it is based most directly on acceptability judgments, on judgments about how various expressions would be interpreted in various contexts, and so on. There is a lively ongoing debate about whether the resulting linguistic generalizations, such as the ones in standard linguistics texts, are empirically stable and well-supported, with some evidence that they are [38, 37]. But this debate is not mentioned by Stokhof and van Lambalgen.

Another motivation for Stokhof and van Lambalgen seems to be a concern that there is something suspect or something "ontologically" unnatural about a psychological mechanism that defines an infinite language, that is, intuitively, a language with patterns that are not bound to some fixed finite lengths. This worry could have been raised directly too, but again it seems insubstantial, and any worries about this strategy apply not only to linguistics but to the vast range of computational descriptions in the sciences, from models of cell-signalling, to models of human problem solving, to the accounts of our own artifacts – calculators and computers.[4]

2 Language processing, neurophysiological realization

Chomsky emphasizes that, "When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed, in some practical or efficient way, to construct such a derivation" [9, p.9]. The empirical basis of linguistic theory does not derive from performance models, though of course we expect compatibility. To obtain parsimonious, compatible psychological, neurological and linguistic models, a natural methodological stance is to look for how the linguistic model is realized in the psychological and neurological models.[5] Models of the recognition of fluent speech, or of lexical decision latencies, for example, typically make various assumptions about how grammatical analysis is influenced by frequency and contextual effects. As we know from computer engineering, it is not quite clear what the relevant notion of "realizing" one model with another should be. Even in the artificial case of "compiling" one program into another, there are questions about what should count as a relevantly faithful realization [5, 32]. For computation in naturally occurring systems, the questions are even more challenging. A great deal of work is devoted to

[4] Standard introductions to computation sometimes briefly remark the step of setting aside memory limitations: "The computer itself can be viewed as a finite state system, although doing so turns out to be not as useful as one would like" [19, p.14]. This does not make students terribly confused about how to use infinite models to describe finite systems, as we do everyday when we think of our calculators as multiplying integers, for example. Connectionist systems and other sorts of neural models are also compared by considering their computational power when temporal and hardware limitations are set aside, as in [33, p.118], [34], for example. If this is a fundamental ontology-changing step, it is not one that necessarily creates great difficulties for understanding the capabilities of finitely bounded physical realizations of those models.

[5] Cf., e.g., [3, pp.75ff] on psycholinguistic models, and cf. e.g. [25, 26] for similar perspectives in the perceptual domain of vision.

understanding neural computation, from speculations about how linguistic analysis could be implemented in simple neural networks [36, 35, 14, directly inspired by generative grammar, for example] and models of how language processing seems to be localized in the brain [1, 7, 41, also grammar inspired] to much simpler questions about how simple neural networks could realize "perceptual magnet" effects [17, 11, for example] and much more complete models of how frogs could recognize the position of prey [12, 13, for example]. Although it may seem that all of our original questions about linguistic performance models remain open, I think that any reasonable assessment of the field would conclude that enormous progress has been made since computational models of various kinds were first emerging in the 1950's, including much directly based on generative grammar.

3 Diversity and consensus

Stokhof and van Lambalgen suggest that "uniformity and consensus ... have disappeared: there is enormous variety of approaches, theoretical models, methodologies, and even with regard to the goals of linguistics and its very object of investigation there are fundamental differences of opinion." A diversity of approaches, models and methodologies is found in all the healthy sciences, but what Stokhof and van Lambalgen seem to be worried about is really that there seem to be different, incompatible treatments of essentially the same phenomena, grounded in undefended methodological assumptions, perhaps "conceptualizing" the phenomena differently. But this is not what we find in syntactic theory or other mainstream parts of linguistics. It might appear that there are great disagreements, but I think the differences are built on a dominant and very wide-ranging consensus.

One important aspect of this consensus was noted by Joshi in 1985 [20] and then extended by many others. Considering the independently developed syntactic frameworks of categorial grammar proposed by Steedman, tree adjoining grammar proposed by Joshi, and a version of generalized phrase structure grammar proposed by Pollard, it was discovered that these apparently very different frameworks are expressively exactly equivalent [42]. Not only that, but the proofs of their weak equivalence are rather easy and constructive, providing recipes for converting derivations in one framework into derivations in the other frameworks. Some time later, it was discovered that a formalization of Chomskian syntax is also exactly equivalent to a variant of tree adjoining grammar and also to a version of generalized phrase structure grammar, and of dependency grammar, and of abstract categorial grammars [27, 18, 28, 6, 30, 31, 21]. Again, this was established by rather simple, constructive proofs. And even more recent work shows that a wide range of constraint-based approaches to grammar define exactly the same class [22, 16, 15]. These remarkable results about independently proposed and superficially very diverse formalisms have sharpened the questions about whether the consensus is right: Are these grammars really adequate? This has led to an ongoing debate that is considerably more interesting (and more collegial) than previous disputes about the adequacy of different linguistic frameworks [2, 29, 39, 4, 23].

4 Conclusion, and the decreasing returns of meta meta...

Stokhof and van Lambalgen offer little to support their perspective, and I would be surprised if this short response offers anything that persuades them either. I see an expanding, ever-shifting, ramifying boom where they see a bust. "Meta-science" of this sort – opinionated, sketchy, informal assessments of whole research traditions or even whole sciences – are rather hard to avoid in the scientific literature, but meta-science is almost always very much less interesting than science. In the present case, especially, it is difficult to have much of interest to say about matters as ill-defined as the linguistic traditions. One always

wants to say: OK, granting your assumptions, let's see what you can do! Unfortunately the present work, a comment on a meta-linguistic work, surely counts as meta-meta-linguistics, which can only be worse! But we can summarize very briefly. How should a student decide among competing linguistic traditions? Perhaps everyone can agree on at least this: the traditions should be judged by their achievements. The best students use all means at their disposal to find the most beautiful and empirically rich theories, those most likely to lead to real insights, those with the best-informed and clearest-thinking teachers. With this highly non-deterministic strategy, it is no surprise to find good students pursuing a variety of approaches – a welcome result. Aiming for some further consensus at these meta-levels, consider these three optimistic notes about the situation. *First*, when linguistic traditions are poorly defined, they can be molded to accommodate almost anything, without seeming to lose their identity, and in that case it is not easy to find principled grounds for worrying too much about any particular choice. (This same lack of definition explains in large part the sterility of most meta-disputes, and the inappropriateness of excessive parochialism.) *Second*, pushed by basic facts that even opposing traditions must agree on, theories that develop in steadfast opposition to each other may end up becoming notational variants, or nearly so, as mentioned in §3 above. *Third*, as at any intellectual feast, it is nice to choose at least one thing that is almost sure to be good, but each of us is free to choose more than one thing, perhaps with attention to making choices that are likely to complement each other nicely.

References

[1] Bachrach, A. *Imaging Neural Correlates of Syntactic Complexity in a Naturalistic Context*. PhD thesis, Massachusetts Institute of Technology, 2008.

[2] Becker, T., Joshi, A. K., and Rambow, O. Long-distance scrambling and tree adjoining grammars. In *Proceedings of the Fifth Conference of the European Association for Computational Linguistics* (Morristown, New Jersey, 1991), Association for Computational Linguistics, pp. 21–26.

[3] Berwick, R. C., and Weinberg, A. S. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. MIT Press, Cambridge, Massachusetts, 1984.

[4] Bhatt, R., and Joshi, A. Semilinearity is a syntactic invariant: a reply to Michaelis and Kracht. *Linguistic Inquiry 35* (2004), 683–692.

[5] Blass, A., Dershowitz, N., and Gurevich, Y. When are two algorithms the same? University of Michigan, Tel Aviv University, and Microsoft Research. http://arxiv.org/abs/0811.0811, 2008.

[6] Boston, M. F., Hale, J. T., and Kuhlmann, M. Dependency structures derived from minimalist grammars. In *Proceedings of Mathematics of Language 11* (2009), pp. 11–20.

[7] Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., and Pylkkänen, L. Syntactic structure building in the anterior temporal lobe during natural story listening. *Forthcoming* (2010).

[8] Cartwright, N. Do the laws of physics state the facts? *Pacific Philosophical Quarterly 61* (1980), 75–84.

[9] Chomsky, N. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Massachusetts, 1965.

[10] Earman, J., Glymour, C., and Mitchell, S. *Ceteris Paribus Laws*. Springer, Berlin, 1991.

[11] Feldman, N. H., and Griffiths, T. L. A rational account of the perceptual magnet effect. In *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society* (2007).

[12] Franosch, J.-M. P., Lingenheil, M., and van Hemmen, J. L. How a frog can learn what is where in the dark. *Physical Review Letters 95*, 7 (2005), 078106.

[13] Franosch, J.-M. P., Sobotka, M. C., Elepfandt, A., and van Hemmen, J. L. Minimal model of prey localization through the lateral-line system. *Physical Review Letters 91*, 15 (2003), 158101.

[14] Gerth, S., and beim Graben, P. Unifying syntactic theory and sentence processing difficulty through a connectionist minimalist parser. *Cognitive Neurodynamics 3*, 4 (2009), 297–316.

[15] Graf, T. Global optimality in endocentric optimality systems. UCLA, forthcoming, 2010.

[16] Graf, T. Reference-set constraints as linear tree transductions via controlled optimality systems. In *Proceedings of the Conference on Formal Grammar, ESSLLI'10* (2010).

[17] Guenther, F. H., and Gjaja, M. N. The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America 100* (1996), 1111–1121.

[18] Harkema, H. A characterization of minimalist languages. In *Logical Aspects of Computational Linguistics* (NY, 2001), P. de Groote, G. Morrill, and C. Retoré, Eds., Lecture Notes in Artificial Intelligence, No. 2099, Springer, pp. 193–211.

[19] Hopcroft, J. E., and Ullman, J. D. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, Reading, Massachusetts, 1979.

[20] Joshi, A. How much context-sensitivity is necessary for characterizing structural descriptions. In *Natural Language Processing: Theoretical, Computational and Psychological Perspectives*, D. Dowty, L. Karttunen, and A. Zwicky, Eds. Cambridge University Press, NY, 1985, pp. 206–250.

[21] Kanazawa, M. Multiple context-free languages and non-duplicating macro languages. In *Proceedings, Workshop on Multiple Context Free Grammars and Related Formalisms, Tokyo* (2010).

[22] Kepser, S., and Mönnich, U. Properties of linear context free tree languages with an application to optimality theory. *Theoretical Computer Science 354* (2006), 82–97.

[23] Kobele, G. M. *Generating Copies: An Investigation into Structural Identity in Language and Grammar*. PhD thesis, UCLA, 2006.

[24] Lange, M. Natural laws and the problem of provisos. *Erkenntnis 38* (1993), 233–248.

[25] Marr, D. Artificial Intelligence—A Personal View. *Artificial Intelligence 9* (1977), 37–48.

[26] Marr, D. *Vision*. Freeman, San Francisco, 1982.

[27] Michaelis, J. Derivational minimalism is mildly context-sensitive. In *Proceedings, Logical Aspects of Computational Linguistics, LACL'98* (NY, 1998), Springer, pp. 179–198.

[28] Michaelis, J. Transforming linear context free rewriting systems into minimalist grammars. In *Logical Aspects of Computational Linguistics* (NY, 2001), P. de Groote, G. Morrill, and C. Retoré, Eds., Lecture Notes in Artificial Intelligence, No. 2099, Springer, pp. 228–244.

[29] Michaelis, J., and Kracht, M. Semilinearity as a syntactic invariant. In *Logical Aspects of Computational Linguistics* (NY, 1997), C. Retoré, Ed., Springer-Verlag (Lecture Notes in Computer Science 1328), pp. 37–40.

[30] Mönnich, U. Minimalist syntax, multiple regular tree grammars, and direction preserving tree transductions. In *Model Theoretic Syntax at 10. ESSLLI'07 Workshop Proceedings* (2007), J. Rogers and S. Kepser, Eds.

[31] Mönnich, U. Well-nested tree languages and attributed tree transducers. In *The 10th International Conference on Tree Adjoining Grammars and Related Formalisms TAG+10* (2010).

[32] Moschovakis, Y. N. What is an algorithm? In *Mathematics unlimited – 2001 and beyond*, B. Engquist and W. Schmid, Eds. Springer, NY, 2001, pp. 919–936.

[33] Rumelhart, D. E., and McClelland, J. L. PDP models and general issues in cognitive science. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1*. MIT Press, Cambridge, MA, USA, 1986.

[34] Siegelmann, H. T., and Sontag, E. D. Turing computability with neural nets. *Applied Mathematics Letters 4*, 6 (1991), 77–80.

[35] Smolensky, P. Grammar based connectionist approaches to language. *Cognitive Science 23*, 4 (1999), 589–613.

[36] Smolensky, P., Goldrick, M., and Mathis, D. Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Forthcoming* (2011). Johns Hopkins University.

[37] Sprouse, J. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods 43* (2011), 155–167.

[38] Sprouse, J., and Almeida, D. A formal experimental investigation of the empirical foundation of generative syntactic theory. *Forthcoming* (2010). University of California, http://www.socsci.uci.edu/∼jsprouse/.

[39] Stabler, E. P. Varieties of crossing dependencies: Structure dependence and mild context sensitivity. *Cognitive Science 93*, 5 (2004), 699–720.

[40] Stokhof, M., and van Lambalgen, M. Abstractions and idealisations: The construction of modern linguistics. http://staff.science.uva.nl/∼stokhof/papers.html. Publication forthcoming, 2010.

[41] VanWagenen, S., Brennan, J., and Stabler, E. P. Evaluating parsing strategies in sentence processing. In *Proceedings of the CUNY Sentence Processing Conference* (2011).

[42] Vijay-Shanker, K., and Weir, D. The equivalence of four extensions of context free grammar formalisms. *Mathematical Systems Theory 27* (1994), 511–545.