Minimalist parsing ed. by Robert C. Berwick and Edward P. Stabler (review)


LANGUAGE
A JOURNAL OF THE LINGUISTIC SOCIETY OF AMERICA

Cristiano Chesi

Language, Volume 98, Number 4, December 2022, pp. 852-858 (Review)

Published by Linguistic Society of America
DOI: *https://doi.org/10.1353/lan.2022.0024*

REVIEW

Minimalist parsing. Ed. by Robert C. Berwick and Edward P. Stabler. Oxford: Oxford University Press, 2019. Pp. 192. ISBN 9780198795094. $45.

Reviewed by Cristiano Chesi, *University School for Advanced Studies (IUSS) Pavia*

The recent popularity of very large corpora and the availability of sophisticated language modeling approaches (e.g. GPT-3; Brown et al. 2020) have revitalized interest in natural language processing (NLP), but have also led to the almost total eclipse of the structuralist approach to parsing and (more generally) of the considerable achievements obtained in formal language studies. In this context, the minimalist program (Chomsky 1995), which is discussed in this book and in which the editors, Robert C. Berwick and Edward P. Stabler, played a remarkable leading role,[1] may sometimes be disregarded. Yet this enterprise deserves serious attention from the following audiences: (i) those who do not want to give up on explanatorily adequate models (in Chomsky's 1965 sense; see also Chomsky 2020), (ii) those who still prefer explicit symbolic formalisms (HPSG, LFG, CCG, TAG, among others) to subsymbolic ones (e.g. recurrent neural networks with various levels of sophistication), and (iii) those who approach language studies from a purely statistical perspective (including machine learning and attention-based networks enthusiasts) that does not sufficiently account for certain intricacies of language.

This collection includes an introduction and six contributions to a 2015 workshop held at MIT on 'Parsing approaches under the minimalist umbrella'. The first three contributions focus on various aspects of minimalist formalization (Ch. 1 by Robert C. Berwick and Edward P. Stabler, Ch. 2 by Sandiway Fong and Jason Ginsburg, and Ch. 3 by Ginsburg and Fong). The book then turns to specific linguistic phenomena: parsing-prosody interaction (Ch. 4 by Kristine M. Yu) and ellipsis resolution (Ch. 5 by Gregory M. Kobele). The final two contributions discuss efficient (and cognitively plausible) parsing strategies (Ch. 6 by Tim Hunter) and propose an explicit 'linking theory' between minimalist parsing predictions and neurometabolic activity recorded during an ecological story-listening task (Ch. 7 by Jixing Li and John Hale).

Ch. 1 (by Berwick and Stabler) is more than an introduction to the forthcoming contributions: it sets the stage for the minimalist enterprise by proposing a minimum common formal denominator for otherwise heterogeneous approaches. It also states the major research goals that this volume aims to address: first, evaluating the actual computational implications and the consistency/correctness of the (often) informally formulated linguistic intuitions under a coherent framework; and second, evaluating how performance evidence (from both the psycholinguistic and the neurolinguistic perspective) can fit with specific complexity metrics built on the proposed parsing models. It is clear that this collection is greatly indebted to the formalism of Stabler 2011, which is efficiently presented in this chapter. The introduction includes a specification of lexical items with their features: morphosyntactic categories and their selection/licensing requirements for 'destructive' feature-driven structure-building operations. For instance, from Stabler's perspective, Merge(α, β) will be successful if and only if either α or β is such that one item selects the morphosyntactic categories of the other, in the end projecting over the one selected. Given the lexical items in 1, in which the double colon associates a specific phonetic realization of the lexical item (to the left) with an ordered set of features (after the double colon), we can define Merge as exemplified in 2.[2]

[1] Berwick significantly contributed to the understanding of the relation between computational complexity and processing difficulty (Barton, Berwick, & Ristad 1987) and, more recently, to the so-called biolinguistics perspective (Friederici et al. 2017), while Stabler formalized the first complete minimalist grammar (Stabler 1997), which remains the reference model for all of the extended proposals discussed in this collection and in the literature in general.

[2] =X indicates the selection of an X category, and ε is a phonetically empty item; distinct words are separated by commas.

(1) Praises :: =D =D V, Bo :: D, Cal :: D, which :: =N D -wh, student :: N, ε :: =V C, ε :: =V +wh C

(2) i. Merge([praises $_{=D\ =D\ V}$], [Bo $_{D}$]) = [praises $_{=D\ V}$ [praises $_{\cancel{=D}\ =D\ V}$ Bo $_{\cancel{D}}$]]

ii. Merge([Cal $_{D}$], [praises $_{=D\ V}$ …]) = [praises $_{V}$ [Cal $_{\cancel{D}}$ [praises $_{\cancel{=D}\ V}$ [praises $_{\cancel{=D}\ =D\ V}$ Bo $_{\cancel{D}}$]]]]

iii. Merge([ε $_{=V\ C}$], [praises $_{V}$ …]) = ε $_{C}$ [ε $_{\cancel{=V}\ C}$ [praises $_{\cancel{V}}$ …]]

By extending the feature set categories with +Y (licensors) and −Y (licensees), we can also define internal Merge, namely Move, as in 3.[3]

(3) Move(ε $_{\cancel{=V}\ +wh\ C}$ [praises $_{\cancel{=D}\ \cancel{V}}$ [Cal $_{\cancel{D}}$] [praises $_{\cancel{=D}\ =D\ V}$ [which $_{\cancel{=N}\ \cancel{D}\ -wh}$ [$_{\cancel{N}}$ student]]] = [ε $_{\cancel{+wh}\ C}$ [which $_{\cancel{-wh}}$ [$_{\cancel{N}}$ student]]$_i$ [ε $_{\cancel{=V}\ +wh\ C}$ [praises $_{\cancel{=D}\ \cancel{V}}$ [Cal $_{\cancel{D}}$] [praises $_{\cancel{=D}\ =D\ V}$ t_i]]]]]

The logic behind the two operations is the same: two features must match to obtain a phrase, relying on either external Merge (a lexical item taken from the lexicon) or internal Merge (i.e. move an item already merged in the structure and remerged at the edge of the phrase built so far). Various considerations extend or restrict this logic in the book by adopting a ‘parsing’ perspective.

At first, Berwick and Stabler clarify that the notion of ‘parsing’ adopted here includes both the classic hierarchical structure-building procedure and a more conventional generation procedure. The first approach is based on an algorithm that has a grammar and a sequence of words as an input, for which at least one phrase structure must be returned if the sentence is grammatical. The second takes the grammar and derives the final word order for a given array of lexical items by rearranging them according to the structure-building operations Merge and Move. A fundamental difference between the two approaches is that under the second interpretation we can directly use standard Merge and Move operations, as defined in mainstream minimalism: that is, as brick-over-brick operations (i.e. ‘from-bottom-to-top’; Phillips 1996), usually leading to a right-to-left generation (e.g. [praises Bo], [Cal [praises Bo]]). Meanwhile, standard parsing procedures receive the word sequence as input (i.e. <Cal, praises, Bo>) and must figure out the morphosyntactic features associated with these items, as well as their hierarchical structure, ‘from-left-to-right’, possibly incrementally, to aim at psycho/neurolinguistic plausibility.

Fong and Ginsburg (Ch. 2 and Ch. 3) pursue the ‘generation’ perspective and provide an orthodox computational implementation of the minimalist framework as discussed in Chomsky 2001. Their goal is to frame various notions in a computationally efficient architecture that is able to account for a relevant set of empirical phenomena. In these two chapters, the reader will find a sober discussion of the implications of various implementations of the possible logical options available. For instance, ‘driven’ vs. ‘free’ structure-building operations are compared: ‘free’ operations simply do not pose constraints on phrase structure building, and the results of a wrong phrasal hypothesis must be discharged using filters independent from the Merge and Move operations. A relevant concern is related to the complexity of free Move/internal Merge; in the end the authors favor a featural-approval-driven operation, which is a probe-goal dependency. Featureless theories are apparently ‘simpler’ in terms of assumptions, and this might be coherent with the evolvability and learnability perspective (Chomsky 2020). However, filtering ill-formed sentences would be costly, even if another module would deal with this (i.e. the CONCEPTUAL-INTENTIONAL module under the LABELING approach proposed by Chomsky 2013). Economy conditions are also considered (such as the MINIMAL LINK CONDITION, ANTILOCALITY, T-to-C MOVEMENT; Pesetsky & Torrego 2001). The conclusion is that removing feature-checking necessities will dispense with most assumptions about the biological component but will make the learning problem much harder (and probably intractable). The concrete proposal consists of a full-fledged procedure (a ‘machine’) that uses a memory stack to keep track of various features to be valued within the current syntactic object, and a set of actions transforming a syntactic object into another syntactic object according to its edge and the status of the stack in a strictly ‘from-bottom-to-top’ perspective. This procedure is nearly determinist and ‘minimal’, in the sense that it reduces as much as possible the memory burden by adopting a ‘quick evaluate and forget’ ap-

[3] The derivation here is just a simplified example; aux-subject inversion and V-to-T movement can readily be implemented.

proach (which is incompatible with, for instance, certain implementations of the labeling algorithm; Bošković 2016).

A naive reader might consider the task of building a computer program with the intent of showing how the derivation of a specific sentence unfolds, step-by-step, according to specific assumptions to not be worth the effort, since it provides us with no new information. This might be true, but only in a grammatical framework that is fully formalized. This is not the case of minimalism: the current status of each minimalist assumption is sometimes vague and inconsistent, or (at best) in constant evolution. This gives computer scientists, who would otherwise simply need to implement a consistent full-fledged grammatical formalism in parsing, a hard time. Fong and Ginsburg's efforts are then crucial for prompting a discussion about which explicit assumptions are missing in the current linguistic theorizing and what the impact of specific implementations would be.

Especially enlightening is the discussion of the inefficiency of free vs. feature-driven operations (Ch. 2) and the necessity of feature unification to deal with Multiple Agree cases (Ch. 3). These assumptions successfully deal with the inherent intricacies of syntactic phenomena, such as expletive insertion or the *that*-trace effect. Every theory that is discussed in the literature presents computational challenges that need to be faced. Fong and Ginsburg convincingly propose their own implementation that, in the end, builds derivations that are in line with the current understanding of the phenomena discussed, without simplifications. This is especially noteworthy given that a common tendency in NLP these days is oversimplification. Minimalism poses a lower boundary on this trend, and Fong and Ginsburg's work reminds us clearly about the subtle syntactic contrasts that we need to consider if we aim for a grammatical account that is, at minimum, descriptively adequate. Among the crucial differences between 'simpler' minimalist approaches (Stabler 1997) is the distinction between two kinds of Merge: both are binary functions, but one produces unordered sets, and the other produces ordered sets (Pair Merge, as in Chomsky's recent work). The authors' attempt to integrate various ideas (Pesetsky & Torrego 2001, Gallego 2006, Sobin 2014, among others) is successful and produces the expected analysis with an acceptable computational cost, which results from limiting nondeterminism in pursuing as far as possible the 'quickly evaluate and forget' maxim.

In the end, the reader might have expected some parsing-oriented considerations, but the authors decided to maintain a rigid bottom-to-top derivation. Therefore, the impact of the many technical solutions adopted to derive the correct structural description from a 'classic' parsing perspective remains a bit obscure.

Yu (Ch. 4) focuses on the complex relationship between prosodic structure and syntactic parsing. This chapter does not answer the question of how prosodic cues might help in disentangling ambiguous plausible syntactic structures, but it does provide a clear background on prosodically informed syntactic parsers[4] and a full-fledged implementation for simple ergative-absolutive configurations in transitive and intransitive constructions in Samoan (a Polynesian language). In this chapter, Yu first introduces the computational reader, who might not be entirely familiar with phonological suprasegmental analyses, to prosodic trees, both highlighting classic concerns about mismatches between syntactic and prosodic boundaries and providing insightful intuitions about how potential mismatches may in fact just be apparent mismatches (Wagner 2010). The author pursues the explanation of these apparent contrasts within the MATCH THEORY framework (Selkirk 2011): prosodic structure reflects syntactic structure, but the two are distinct. The prosodic component relies on the notion of markedness to penalize those solutions that are allowed but that violate some syntax-prosody Match constraint (e.g. MATCHPHRASE: Phonological phrases correspond to maximal projections in syntax). An OPTIMALITY THEORY-like approach is then needed to rank the plausible prosodic trees. Various assumptions are needed to deliver a complete proof of concept: a specific implementation of the lexicon and derivation, together with a graphical implementation of the 'augmented derivation tree' (Graf 2013 is valuable here), one specific parsing algorithm (which in this case is a bottom-up parser; Harkema 2001, Stabler 2013), and one complete procedure to

[4] Table 4.1 is a very handy map to the relevant recent literature in this field.

compute the syntax-prosody match violations and to rank the plausible solutions. The emergent complexity of the whole enterprise might, again, discourage the reader from trying to put all of the pieces together, but the broad perspective gained is worth the effort in the end.

In the fifth chapter, Kobele addresses the efficiency problem of retrieving what is missing (or reduced) but necessary for a correct sentence interpretation once the structural description of a given sentence is retrieved by the parser. The case study is ellipsis, and the solution provided builds on an adaptation of Stabler and Keenan's (2003) 'chain-based' version of minimalist grammars (MGs) under the cyclic spell-out perspective (Chomsky 2000), as discussed in Kobele 2015. The necessary ingredients of this analysis are as follows: (i) a full syntactic structure of the elided constituent is 'in situ', though phonetically null, and (ii) its absence is obtained by applying two constrained operations (i.e. simultaneous substitution and phonological deletion). Much of the effort is devoted to the definition of the contextually salient material that licenses such a transformation-and-deletion operation, that is, ellipsis. The algorithm is compatible with all possible chart-parsing algorithms discussed in Harkema 2001 under the reasonable assumption that only a fixed, finite number of possible (nonrecursive, but see Tomioka 2008) ellipsis operations are allowed, and the problem then becomes efficiently solvable. Kobele attains this by adopting the clever assumption that only maximal projections can be elided: that is, in MGs, any item that exhausts its featural make-up after a finite number of Merge/Move operations can be elided. That each ellipsis must be associated with a unique preceding lexical item becomes a natural assumption, and we then conclude that the number of possible antecedents is bounded with respect to the size of the sentence(s) expressed in terms of lexical items merged so far.

In the end, the solution is clever and sound, but the reader is left with few details about the actual parsing strategies to be adopted in order to efficiently retrieve the relevant structure in cases of ellipsis without overgeneralizing on ill-formed constituents (*_Adam eats_ [_a sandwich_]$_i$ _and Bea eats_ e_i). Some considerations of a minimal parameterized approach to capturing relevant crosslinguistic options (e.g. English ellipsis vs. Romance pronominalization) would have been a useful addition here. Readers would probably need to browse the relevant work cited in the chapter, especially Kobele 2015, to get a clear picture.

Other kinds of nonlocal dependencies are discussed in Hunter's chapter (Ch. 6): classic filler-gap dependencies of the wh-kind are here approached assuming that (i) the (active) filler can be processed before the relevant gap position is postulated, and (ii) a top-down parser with a left-corner filtering strategy is sufficient for the purpose. The parsing strategy chosen is a readaptation of Stabler's (2013) approach. The intuition is that the left-most expansion of a given category must be the result of a bottom-up (category) shift, while its expectation is top-down driven (as in context-free grammars; CFGs). In both Stabler's and Hunter's approaches (and also in Kobele's), such structure-building operations cannot simply be the reverse of minimalist Merge and Move, as conceived, for instance, in Chomsky 2013. Being feature-driven operations, they must consider (at least) a pair of options for each instance of structure building; that is, minimally, either the probe precedes the goal, or the other way around: $\langle \alpha_{=/+X}, \beta_{(-)X} \rangle$, yielding α, or $\langle \alpha_{(-)X}, \beta_{=/+X} \rangle$, yielding β. The CFG left-corner strategy is naturally translated into a shifting rule (which consumes an input word and substitutes it in the stack with its 'category', that is, a list of features) and a scanning rule (which consumes a predicted category from the stack). Their interleaving with predict and connect rules does the rest (the first expanding a category given its left corner, the second substituting a fulfilled categorial expectation with another prediction based on the Merge hypothesis). The basic derivations of simple sentences with and without movement proceed smoothly, but complications arise in extending this approach to smuggling (this is the case of [α [… [β]]] in a derivation in which first α moves, then β moves, such as in a passive derivation, as proposed by Collins 2005) and remnant movement (which is the reverse case: in [α [… [β]]], first β, then α moves). These cases, as well as (successive) cyclic movement, are not just exotic puzzles to be solved but are relevant possibilities that every parsing approach should consider.[5] Even though not all of the puzzles are read-

[5] Smuggling is becoming a very popular analysis in many derivations that are apparently able to avoid intervention effects. See postverbal subject constructions in Italian, as analyzed in Belletti & Chesi 2014.

ily solvable under the presented perspective,[6] the discussion in this chapter is praiseworthy because it indicates the author's awareness of the subtle aspects that are tentatively modeled.

Another relevant aspect mentioned in this chapter is the support found in the psycholinguistics literature for some of the implemented alternatives, such as the 'hyperactive filler' hypothesis (Omaki et al. 2015): that is, the parser posits a gap as soon as it can, whenever it can (possibly in illegitimate nested positions, in case the structure would be salvaged by means of a parasitic + licensed gaps configuration; Bianchi & Chesi 2006).

This cognitive plausibility (broadly speaking) is the main concern of Li and Hale's chapter (Ch. 7), which concludes the book with a very elegant incursion into performance tasks: how does the predicted structural processing impact neurometabolic activity that occurs while listening to a story? The authors assess the improvement of a regression model aimed at predicting the recorded blood-oxygen-level-dependent (BOLD) signal variations in four specific brain regions of interest (the left anterior temporal (LATL), right anterior temporal (RATL), and left posterior temporal (LPTL) lobes, and the left inferior frontal gyrus (LIFG)) that were obtained while participants listened to the first chapter of *Alice in Wonderland* (Brennan et al. 2016). Various predictors are considered, including both purely linear (N-gram-based) surprisal models (Hale 2016) and others based on a combination of grammatical models and parsing strategies. Vectorial representations of the meaning of the words are also included as regressors (as a measure of lexical-semantic coherence).

Two results are especially worth stressing: first, various parsing algorithms (mainly top-down and bottom-up) correlate pretty well under the same grammatical formalism (i.e. CFG or MG); second, statistical measures based on surprisal (obtained both by using simple N-grams and by adopting the EarleyX algorithm, a top-down algorithm, to CFG; Stolcke 1995) significantly contribute to the prediction of the activation patterns in all areas of interest. Surprisingly, structural-distance metrics, based on both CFG and MG, correlate with the activation pattern only in the LPTL lobe and not in Broca's-related regions, like LIFG. The authors suggest that this might be related to the limited effort that is required in parsing grammatical sentences. This conclusion is supported by the fact that evidence for Broca's area activation is mainly obtained through the elicitation of differential activations triggered by specific violations as compared to correct structures (i.e. Musso et al. 2003); that is, increased activity is recorded in these areas where morphosyntactic violations are observed (but see Pallier, Devauchelle, & Dehaene 2011, among others, for a different, equally supportive approach). Meanwhile, as discussed by the authors, it is practically impossible to disentangle morphosyntactic, semantic, and pragmatic factors (if these labels still make any sense) in a naturalistic task such as story-listening. The lack of correlation might then not be so surprising, and more minimal manipulations of the 'story-listening' materials might be imagined for obtaining meaningful contrasts.

Finally, the authors clearly show how important it is to explicitly indicate a 'linking theory' that maps a recordable behavior (self-paced reading, eye-tracking, fMRI, etc.) with a specific theory: the adoption of a specific formalism (e.g. CFG vs. MG) is insufficient, so a parsing strategy must also be formulated (e.g. adopting a shift-reduced bottom-up algorithm or an Earley-like top-down approach). The fact that this choice does not impact the regression-fitting results might indicate that another assumption is misleading: assuming a perfect oracle to make the correct/most likely choice in any case might in fact obfuscate the (possibly small) contribution that structural or lexical ambiguity might add as a term of structural complexity.

All in all, this collection is to be commended for going into concrete (parsing and empirical) problems that need precise, explicit modeling: that is, from the minimalist perspective, a useful form of reductionism enabling us to better define the relevant factors that characterize our language processing. The minimalist approach strongly simplified the generativist framework by

[6] Just to mention two of them: criterial freezing (Rizzi & Shlonsky 2008) presents a problem for the smuggling analysis, which seems, however, necessary if reconstruction is admitted (Bianchi & Chesi 2014); and successive cyclic movement does not seem to 'consume' features in a relevant sense as proposed in §6.3.3 (cf. Chesi 2015).

both retaining the PRINCIPLES-AND-PARAMETERS intuition and introducing basic structure-building operations (essentially Merge) that are computationally and biolinguistically appealing. However, the vagueness (or stipulation) of some feature-driven solutions has turned off many computational scholars' interest in this framework. Discussions like the ones in this book are the only way that I foresee to reconcile this audience with some really powerful intuitions.

REFERENCES

Barton, G. Edward; Robert C. Berwick; and Eric Sven Ristad. 1987. *Computational complexity and natural language: Computational models of cognition and perception*. Cambridge, MA: MIT Press.

Belletti, Adriana, and Cristiano Chesi. 2014. A syntactic approach toward the interpretation of some distributional frequencies: Comparing relative clauses in Italian corpora and in elicited production. *Rivista di Grammatica Generativa* 36.1–28. Online: https://lingbuzz.com/j/rgg/2014/belletti+chesi_rivista_2014.pdf.

Bianchi, Valentina, and Cristiano Chesi. 2006. Phases, left branch islands, and computational nesting. *University of Pennsylvania Working Papers in Linguistics (Proceedings of the 29th annual Penn Linguistics Colloquium)* 12(1).15–28. Online: https://repository.upenn.edu/pwpl/vol12/iss1/3.

Bianchi, Valentina, and Cristiano Chesi. 2014. Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry* 45.525–69. DOI: 10.1162/LING_a_00166.

Bošković, Željko. 2016. On the timing of labeling: Deducing Comp-trace effects, the subject condition, the adjunct condition, and tucking in from labeling. *The Linguistic Review* 33.17–66. DOI: 10.1515/tlr-2015-0013.

Brennan, Jonathan R.; Edward P. Stabler; Sarah E. Van Wagenen; Wen-Ming Luh; and John T. Hale. 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language* 157–158.81–94. DOI: 10.1016/j.bandl.2016.04.008.

Brown, Tom B.; Benjamin Mann; Nick Ryder; Melanie Subbiah; Jared Kaplan; Prafulla Dhariwal; Arvind Neelakantan; Pranav Shyam; et al. 2020. Language models are few-shot learners. arXiv:2005.14165 [cs.CL]. Online: http://arxiv.org/abs/2005.14165.

Chesi, Cristiano. 2015. On directionality of phrase structure building. *Journal of Psycholinguistic Research* 44.65–89. DOI: 10.1007/s10936-014-9330-6.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130.33–49. DOI: 10.1016/j.lingua.2012.12.003.

Chomsky, Noam. 2020. The UCLA lectures (April 29–May 2, 2019). Online: https://lingbuzz.net/lingbuzz/005485.

Collins, Chris. 2005. A smuggling approach to raising in English. *Linguistic Inquiry* 36.259–98. DOI: 10.1162/0024389053710701.

Friederici, Angela D.; Noam Chomsky; Robert C. Berwick; Andrea Moro; and Johan J. Bolhuis. 2017. Language, mind and brain. *Nature Human Behaviour* 1.713–22. DOI: 10.1038/s41562-017-0184-4.

Gallego, Ángel J. 2006. T-to-C movement in relative clauses. *Romance languages and linguistic theory 2004: Selected papers from 'Going Romance', Leiden, 9–11 December 2004*, ed. by Jenny Doetjes and Paz González, 143–70. Amsterdam: John Benjamins.

Graf, Thomas. 2013. *Local and transderivational constraints in syntax and semantics*. Los Angeles: University of California, Los Angeles dissertation. Online: https://escholarship.org/uc/item/7nt7m7sb.

Hale, John. 2016. Information-theoretical complexity metrics. *Language and Linguistics Compass* 10.397–412. DOI: 10.1111/lnc3.12196.

Harkema, Hendrik. 2001. *Parsing minimalist languages*. Los Angeles: University of California, Los Angeles dissertation.

Kobele, Gregory M. 2015. LF-copying without LF. *Lingua* 166.236–59. DOI: 10.1016/j.lingua.2014.08.006.

Musso, Mariacristina; Andrea Moro; Volkmar Glauche; Michel Rijntjes; Jürgen Reichenbach; Christian Büchel; and Cornelius Weiller. 2003. Broca's area and the language instinct. *Nature Neuroscience* 6.774–81. DOI: 10.1038/nn1077.

Omaki, Akira; Ellen F. Lau; Imogen Davidson White; Myles L. Dakan; Aaron Apple; and Colin Phillips. 2015. Hyper-active gap filling. *Frontiers in Psychology* 6:00384. DOI: 10.3389/fpsyg.2015.00384.

Pallier, Christophe; Anne-Dominique Devauchelle; and Stanislas Dehaene. 2011. Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences* 108.2522–27. DOI: 10.1073/pnas.1018711108.

Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 355–426. Cambridge, MA: MIT Press.

Phillips, Colin. 1996. *Order and structure*. Cambridge, MA: MIT dissertation. Online: http://www.ai.mit.edu/projects/dm/theses/phillips96.pdf.

Rizzi, Luigi, and Ur Shlonsky. 2008. Strategies of subject extraction. *Interfaces + recursion = Language?: Chomsky's minimalism and the view from syntax-semantics*, ed. by Uli Sauerland and Hans-Martin Gärtner, 115–60. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110207552.115.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. *The handbook of phonological theory*, 2nd edn., ed. by John Goldsmith, Jason Riggle, and Alan C. L. Yu, 435–84. Oxford: Blackwell. DOI: 10.1002/9781444343069.ch14.

Sobin, Nicholas. 2014. Th/Ex, agreement, and case in expletive sentences. *Syntax* 17.385–416. DOI: 10.1111/synt.12021.

Stabler, Edward P. 1997. Derivational minimalism. *Logical aspects of computational linguistics (LACL 1996)*, ed. by Christian Retoré, 68–95. Berlin: Springer. DOI: 10.1007/BFb0052152.

Stabler, Edward P. 2011. Computational perspectives on minimalism. *The Oxford handbook of linguistic minimalism*, ed. by Cedric Boeckx, 617–42. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199549368.013.0027.

Stabler, Edward P. 2013. Two models of minimalist, incremental syntactic analysis. *Topics in Cognitive Science* 5.611–33. DOI: 10.1111/tops.12031.

Stabler, Edward P., and Edward L. Keenan. 2003. Structural similarity within and among languages. *Theoretical Computer Science* 293.345–63. DOI: 10.1016/S0304-3975(01)00351-6.

Stolcke, Andreas. 1995. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics* 21.165–201. Online: https://dl.acm.org/doi/10.5555/211190.211197.

Tomioka, Satoshi. 2008. A step-by-step guide to ellipsis resolution. *Topics in ellipsis*, ed. by Kyle Johnson, 210–28. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511487033.009.

Wagner, Michael. 2010. Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory* 28.183–237. DOI: 10.1007/S11049-009-9086-0.

[cristiano.chesi@iusspavia.it]