COGNITIVE CONSTRAINTS BUILT INTO FORMAL GRAMMARS: IMPLICATIONS FOR LANGUAGE EVOLUTION

CARLOS GÓMEZ-RODRÍGUEZ[1], MORTEN H. CHRISTIANSEN[2,3], and RAMON FERRER-I-CANCHO[*4]

[*]Corresponding Author: rferrericancho@cs.upc.edu
[1]FASTPARSE Lab, LyS Research Group, Departamento de Computación, Universidade da Coruña, CITIC, Campus de Elviña s/n, 15071, A Coruña, Spain
[2]Department of Psychology, Cornell University, Ithaca, NY 14853, USA
[3]Interacting Minds Centre and School of Communication and Culture, Nobelparken, Aarhus University, 8000 Aarhus C, Denmark
[4]Departament de Ciències de la Computació, Universitat Politècnica de Catalunya (UPC), Jordi Girona 1-2, 08034, Barcelona, Catalonia

An often celebrated aspect of human language is its capacity to produce an unbounded number of different sentences (Chomsky, 1965; Miller, 2000). For many centuries, the goal of linguistics has been to capture this capacity by a formal description—a grammar—consisting of a systematic set of rules and/or principles that determine which sentences are part of a given language and which are not (Bod, 2013). Over the years, these formal grammars have taken many forms but common to them all is the assumption that they capture the idealized linguistic competence of a native speaker/hearer, independent of any memory limitations or other non-linguistic cognitive constraints (Chomsky, 1965; Miller, 2000). These abstract formal descriptions have come to play a foundational role in the language sciences (Hauser, Chomsky, & Fitch, 2002; Pinker, 2003). Despite evidence that processing difficulty underpins the unacceptability of certain sentences (Morrill, 2010; Hawkins, 2004), the cognitive independence assumption that is a defining feature of linguistic competence has not been examined in a systematic way using the tools of formal grammar. It is therefore unclear whether these supposedly idealized descriptions of language are free of non-linguistic cognitive constraints, such as memory limitations.

If the cognitive independence assumption should turn out not to hold, then it would have wide-spread theoretical and practical implications for our understanding of human linguistic productivity. It would require a reappraisal of key parts of linguistic theory that hitherto have posed formidable challenges for explanations of language processing, acquisition and evolution (Gold, 1967; Hauser et al., 2002; Pinker, 2003)—pointing to new ways of thinking about language that may simplify the problem space considerably (Christiansen & Chater, 2008; Gómez-Rodríguez & Ferrer-i-Cancho, 2017). Here, we therefore evaluate the cognitive independence assumption using a state-of-the-art grammatical framework, dependency grammar (Nivre, 2005), to search for possible hidden memory constraints in these formal, idealized descriptions of natural language. To delimit the set of possible grammatical descriptions, various classes or sets of syntactic dependency structures have been proposed. These classes can be seen as filters on the set of all the possible syntactic structures. Here, we consider projective structures and various classes of mildly non-projective structures (Gómez-Rodríguez, 2016).

We validate the assumption of independence between grammatical constraints and cognitive limitations in these classes of grammar using the distance between syntactically related words in a dependency tree as a proxy for memory constraints (Liu, Xu, & Liang, 2017). For a given sentence length n, we generate an ensemble of artificial syntactic dependency structures by exhaustive sampling for $n \leq n^* = 10$ and random sampling for $n > n^*$. These artificial syntactic dependency trees are only constrained by the definition of the different classes. They are thus free from any memory constraint other than the ones the different classes of grammars may, perhaps, impose indirectly. Strikingly, while previous work on natural languages has shown that dependency lengths are considerably below what would be expected by a random baseline without memory constraints (Ferrer-i-Cancho, 2004; Ferrer-i-Cancho & Liu, 2014), we still observe a drop in dependency lengths for randomly generated, mildly non-projective structures that supposedly abstract away from cognitive limitations. Our current findings show that memory limitations have permeated current linguistic conceptions of grammar, suggesting that it may not be possible to adequately capture our unbounded capacity for language without incorporating cognitive constraints into the grammar formalism.

Beyond upending longheld assumptions about the nature of human linguistic productivity, our findings also have key implications for debates on how children learn language and how language evolved. Whereas a common assumption in the acquisition literature is that children come to the task of language learning with built-in linguistic constraints on what they learn (Gold, 1967; Pinker, 2003), our results suggest that language-specific constraints may not be needed and instead be replaced by general cognitive constraints (Tomasello, 2005). The strong effects of memory on dependence distance minimization provide further support for the notion that language evolved through processes of cultural evolution shaped by the human brain (Christiansen & Chater, 2008), rather than the biological evolution of language-specific constraints (Pinker, 2003).

Acknowledgements

CGR is suppported by ERC (EU's Horizon 2020; FASTPARSE, 714150), MINECO (TIN2017-85160-C2-1-R, ANSWER-ASAP) and Xunta de Galicia (ED431B 2017/01 and an Oportunius program grant). RFC is supported by MINECO (TIN2017-89244-R) and AGAUR (2017SGR-856, MACDA).

References

Bod, R. (2013). *A new history of the humanities: The search for principles and patterns from antiquity to the present.* Oxford, UK: Oxford University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489 - 558.

Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, *70*, 056135.

Ferrer-i-Cancho, R., & Liu, H. (2014). The risks of mixing dependency lengths from sequences of different length. *Glottotheory*, *5*, 143-155.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*, 447-474.

Gómez-Rodríguez, C. (2016). Restricted non-projectivity: Coverage vs. efficiency. *Computational Linguistics*, *42*(4), 809–817.

Gómez-Rodríguez, C., & Ferrer-i-Cancho, R. (2017). Scarcity of crossing dependencies: a direct outcome of a specific constraint? *Physical Review E*, *96*, 062304.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it and how did it evolve? *Science*, *298*, 1569-1579.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars.* Oxford: Oxford University Press.

Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, *21*, 171-193.

Miller, P. (2000). *Strong generative capacity: The semantics of linguistic formalism.* Cambridge, UK: Cambridge University Press.

Morrill, G. (2010). *Categorial grammar: logical syntax, semantics, and processing.* Oxford: Oxford University Press.

Nivre, J. (2005). *Dependency grammar and dependency parsing* (Tech. Rep. No. MSI 05133). Växjö University, School of Mathematics and Systems Engineering.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (p. 16-37). New York: Oxford University Press.

Tomasello, M. (2005). *Constructing a language. A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.