

On the relation between dependency distance, crossing dependencies, and parsing.

Comment on “Dependency distance: a new perspective on syntactic patterns in natural languages” by Haitao

Liu et al.

Carlos Gómez-Rodríguez

Universidade da Coruña. FASTPARSE Lab, LyS Research Group, Departamento de Computación, Facultade de Informática, Elviña, 15071 A Coruña, Spain

Keywords: dependency distance, natural language parsing, dependency parsing, crossing dependencies, non-projectivity

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



This is the accepted manuscript (final peer-reviewed manuscript) accepted for publication in *Physics of Life Reviews*, and may not reflect subsequent changes resulting from the publishing process such as copy-editing, formatting or pagination. The published journal article can be found at <https://doi.org/10.1016/j.plrev.2017.05.007>.

Liu et al. [1] provide a comprehensive account of research on dependency distance in human languages. While the article is a very rich and useful report on this complex subject, here I will expand on a few specific issues where

Email address: carlos.gomez@udc.es ()

URL: <http://www.grupolys.org/~cgomezr> ()

research in computational linguistics (specifically natural language processing) can inform DDM research, and vice versa. These aspects have not been explored much in [1] or elsewhere, probably due to the little overlap between both research communities, but they may provide interesting insights for improving our understanding of the evolution of human languages, the mechanisms by which the brain processes and understands language, and the construction of effective computer systems to achieve this goal.

1. Crossings, dependency distance and the parallelism between exceptions to projectivity and exceptions to DDM

As mentioned in [1], there is a close relation between DDM and the scarcity of crossing dependencies in natural languages. This low frequency of crossings has long been observed [2], later quantified [3, 4, 5], and recently statistically tested [6], in a wide range of human languages.

It is worth noting, however, that strict projectivity (a prohibition of crossing dependencies) is not an adequate model of the syntax of real sentences. On the one hand, it fails to explain a number of relevant linguistic phenomena present in various languages [7, 8]. On the other hand, an overwhelming majority of syntactic corpora in recent multilingual collections have been observed to contain non-projectivity [4, 5]: crossing dependencies are scarce, but far from absent [6]. For these reasons, while projectivity can be useful from an engineering point of view, in the context of a tradeoff between coverage and efficiency in natural language parsers [5]; taking it for granted when investigating DDM or using it as an assumption of models to explain DDM [9, 10] can lead to a methodological pitfall: the scarcity of crossing dependencies is likely not an independent constraint of language that contributes to DDM, but rather a consequence of it [11, 12].

To account for the limited amount of crossing dependencies that arise in language and to be able to build parsers that can handle non-projective syntactic phenomena in an efficient manner, researchers have explored various classes of so-called mildly non-projective dependency structures [13, 14, 15, 4, 16, 5]. These are sets of trees that allow a limited degree of non-projectivity, permitting crossing dependencies only if they follow certain conditions or patterns. Various such classes have been defined that claim very high coverage over the trees in a variety of corpora, allowing a large majority of the non-projective dependencies that appear in practice. However, the reasons for this success (and especially, the reasons why some of the proposed sets

have more coverage than others) are currently not very well known. Namely, the reasons for the high coverage of each given mildly non-projective class could include either being an adequate description of the particular situations where crossing dependencies can arise, or yielding a large enough class of syntactic trees to provide high coverage by sheer brute force, or a mix of both. The observation that a given class can be more or less adequate depending on the criteria used to annotate the syntactic dependencies [5] suggests at least some influence of the first factor.

The research on syntactic patterns involving long-distance dependencies, reviewed by Liu et al. in Section 5 of [1], could help clarify this question. DDM and the scarcity of crossing dependencies are closely related, as the latter is motivated by short dependency distance [12]. Furthermore, in both cases, there is a predominant trend (the majority of dependencies are short and do not cross) but there are exceptions that escape the trend (long-distance dependencies and crossing dependencies), which are in turn related (longer dependencies are more likely to cross [17]). Liu et al. [1] review some possible reasons for the minority of long dependencies observed in language, and explanations of how they can survive the pressure for DDM. This raises two questions: (1) do these explanations also apply to the presence of crossing dependencies, and are they related with the adequacy of mildly non-projective classes of trees? (e.g., do the most effective such classes work well because they favor crossing dependencies from words at peripheral positions, or structures that can be easily chunked?); and (2) can we in turn re-use some of what we know about long dependencies for crossing dependencies and their parsing, and employ it to define classes of mildly non-projective structures that will more closely adjust to the kinds of non-projectivity found in language? Both questions are interesting avenues for research, and can advance our knowledge both on DDM and natural language parsing.

2. The surprising effectiveness of transition-based parsers and their bias towards DDM

Liu et al. [1] cite some work in computational linguistics that achieved improvements in parsing accuracy by purposefully introducing dependency distance as a constraint in a parser [18]. Additionally, it is worth noting that many state-of-the-art natural language parsers use algorithms with an implicit bias towards short dependency distances, even if they do not introduce it as an explicit restriction.

In particular, a popular framework for dependency parsers is the transition-based (or shift-reduce) approach [19], under which a parser is defined with a non-deterministic state machine, a statistical or machine-learning-based model to score transitions, and a search strategy to obtain the optimal sequence of transitions that will yield a parse. Many, if not most, of the current state-of-the-art parsing systems are based on this framework [20, 21, 22, 23, 24], and all the different algorithm variants that are at its core have in common that they build short dependencies before (and requiring fewer transitions than) long ones, be it because building a long dependency requires removing intervening nodes from a stack (as in the popular arc-standard and arc-eager [19], arc-hybrid [25] or swap [26] algorithms) or because it requires to navigate a list (as in the systems based on the Covington [27] algorithm). This bias towards favoring short dependencies can be part of the reason why these systems are so effective in practice, and is an example of the trend pointed out in [28] by which purely engineering-oriented parsing models are converging with cognitive theories of language understanding, even when they do not have psycholinguistic modeling among their goals.

A quick verification and quantification of the mentioned bias can be undertaken by implementing transition systems and obtaining random trees by taking a random transition at each state. Focusing on sentences of length 20 as an example, the expected mean dependency distance for a uniformly random tree is 7 [29], contrasting with real averages observed in corpora, e.g. 2.52 for Arabic, 3.18 for German, 2.59 for English or 2.51 for Spanish in the sentences of length 20 of the Stanford HamleDT 2.0 treebanks [30]. On the other hand, by simulating 10^5 random parses of sentences of length 20, we obtain an average distance of 2.44 with the arc-standard algorithm, and 2.06 with the arc-eager parser with the tree constraint [31]. These two parsers are projective, but using an algorithm that can generate arbitrary non-projective trees (the swap parser), the obtained average is 2.38, even smaller than for arc-standard. While further investigation is needed that would be outside the scope of this comment, the data seem to suggest that transition-based parsers implicitly favor dependency lengths that are equal, or even smaller, than the natural ones that appear in language as a consequence of DDM; and this could be an important factor in the practical adequacy of these parsers.

Acknowledgements

This research has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 714150 - FASTPARSE), and from the TELEPARES-UDC project (FFI2014-51978-C2-2-R) from MINECO. I thank Ramon Ferrer-i-Cancho for helpful comments.

References

- [1] H. Liu, C. Xu, J. Liang, Dependency distance: A new perspective on syntactic patterns in natural languages, *Physics of Life Reviews* (in this issue). doi:<http://dx.doi.org/10.1016/j.plrev.2017.03.002>.
URL <http://www.sciencedirect.com/science/article/pii/S1571064517300532>
- [2] I. Mel’čuk, *Dependency Syntax: Theory and Practice*, State University of New York Press, 1988.
- [3] J. Havelka, Beyond projectivity: Multilingual evaluation of constraints and measures on non-projective structures, in: *ACL 2007: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 2007, pp. 608–615.
- [4] C. Gómez-Rodríguez, J. Nivre, Divisible transition systems and multi-planar dependency parsing, *Comput. Linguist.* 39 (4) (2013) 799–845. doi:[10.1162/COLI_a_00150](https://doi.org/10.1162/COLI_a_00150).
URL http://dx.doi.org/10.1162/COLI_a_00150
- [5] C. Gómez-Rodríguez, Restricted non-projectivity: Coverage vs. efficiency, *Computational Linguistics* 42 (4) (2016) 809–817. doi:[10.1162/COLI_a_00267](https://doi.org/10.1162/COLI_a_00267).
URL http://dx.doi.org/10.1162/COLI_a_00267
- [6] R. Ferrer-i-Cancho, C. Gómez-Rodríguez, J. L. Esteban, Are crossing dependencies really scarce?, arXiv 1703.08324 [physics.soc-ph].
URL <https://arxiv.org/abs/1703.08324>
- [7] R. Levy, E. Fedorenko, M. Breen, T. Gibson, The processing of extra-posed structures in English, *Cognition* 122 (1) (2012) 12 – 36.

- [8] Y. Versley, Experiments with easy-first nonprojective constituent parsing, in: Proceedings of the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages, Dublin City University, Dublin, Ireland, 2014, pp. 39–53.
URL <http://www.aclweb.org/anthology/W14-6104>
- [9] H. Liu, Dependency distance as a metric of language comprehension difficulty, *Journal of Cognitive Science* 9 (2008) 159–191.
- [10] R. Futrell, K. Mahowald, E. Gibson, Large-scale evidence of dependency length minimization in 37 languages, *Proceedings of the National Academy of Sciences* 112 (33) (2015) 10336–10341. [arXiv:http://www.pnas.org/content/112/33/10336.full.pdf](http://www.pnas.org/content/112/33/10336.full.pdf), doi:10.1073/pnas.1502134112.
URL <http://www.pnas.org/content/112/33/10336.abstract>
- [11] C. Gómez-Rodríguez, R. Ferrer-i-Cancho, The scarcity of crossing dependencies: a direct outcome of a specific constraint?, arXiv 1601.03210 [cs.CL].
URL <http://arxiv.org/abs/1601.03210>
- [12] R. Ferrer-i-Cancho, C. Gómez-Rodríguez, Crossings as a side effect of dependency lengths, *Complexity* 21 (S2) (2016) 320–328. doi:10.1002/cplx.21810.
URL <http://dx.doi.org/10.1002/cplx.21810>
- [13] M. Kuhlmann, J. Nivre, Mildly non-projective dependency structures, in: Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions, 2006, pp. 507–514.
- [14] C. Gómez-Rodríguez, J. A. Carroll, D. J. Weir, Dependency parsing schemata and mildly non-projective dependency parsing, *Computational Linguistics* 37 (3) (2011) 541–586. doi:10.1162/COLI_a_00060.
URL http://dx.doi.org/10.1162/COLI_a_00060
- [15] E. Pitler, S. Kannan, M. Marcus, Dynamic programming for higher order parsing of gap-minding trees, in: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and

Computational Natural Language Learning, Association for Computational Linguistics, Jeju Island, Korea, 2012, pp. 478–488.
URL <http://www.aclweb.org/anthology/D12-1044>

- [16] E. Pitler, S. Kannan, M. Marcus, Finding optimal 1-endpoint-crossing trees, *Transactions of the Association of Computational Linguistics* 1 (2013) 13–24.
URL <http://aclweb.org/anthology/Q13-1002>
- [17] R. Ferrer-i-Cancho, Non-crossing dependencies: least effort, not grammar, in: A. Mehler, A. Lücking, S. Banisch, P. Blanchard, B. Job (Eds.), *Towards a theoretical framework for analyzing complex linguistic networks*, Springer, Berlin, 2016, pp. 203–234.
- [18] J. Eisner, N. A. Smith, Favor short dependencies: Parsing with soft and hard constraints on dependency length, in: H. Bunt, P. Merlo, J. Nivre (Eds.), *Trends in Parsing Technology, Dependency Parsing, Domain Adaptation, and Deep Parsing*, Springer, 2010, pp. 121–150.
doi:10.1007/978-90-481-9352-3_8.
URL http://dx.doi.org/10.1007/978-90-481-9352-3_8
- [19] J. Nivre, Algorithms for Deterministic Incremental Dependency Parsing, *Computational Linguistics* 34 (4) (2008) 513–553.
doi:10.1162/coli.07-056-R1-07-027.
URL <http://www.mitpressjournals.org/doi/abs/10.1162/coli.07-056-R1-07-027>
- [20] D. Chen, C. Manning, A fast and accurate dependency parser using neural networks, in: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 740–750.
URL <http://www.aclweb.org/anthology/D14-1082>
- [21] C. Dyer, M. Ballesteros, W. Ling, A. Matthews, N. A. Smith, Transition-based dependency parsing with stack long short-term memory, in: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Beijing, China, 2015, pp. 334–343.
URL <http://www.aclweb.org/anthology/P15-1033>

- [22] D. Andor, C. Alberti, D. Weiss, A. Severyn, A. Presta, K. Ganchev, S. Petrov, M. Collins, Globally normalized transition-based neural networks, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany, 2016, pp. 2442–2452.
URL <http://www.aclweb.org/anthology/P16-1231>
- [23] E. Kiperwasser, Y. Goldberg, Simple and accurate dependency parsing using bidirectional lstm feature representations, Transactions of the Association for Computational Linguistics 4 (2016) 313–327.
URL <https://transacl.org/ojs/index.php/tacl/article/view/885>
- [24] C. Alberti, D. Andor, I. Bogatyy, M. Collins, D. Gillick, L. Kong, T. Koo, J. Ma, M. Omernick, S. Petrov, C. Thanapirom, Z. Tung, D. Weiss, Syntaxnet models for the conll 2017 shared task, CoRR abs/1703.04929.
URL <http://arxiv.org/abs/1703.04929>
- [25] M. Kuhlmann, C. Gómez-Rodríguez, G. Satta, Dynamic programming algorithms for transition-based dependency parsers, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL 2011), Association for Computational Linguistics, Portland, Oregon, USA, 2011, pp. 673–682.
URL <http://www.aclweb.org/anthology/P11-1068>
- [26] J. Nivre, Non-projective dependency parsing in expected linear time, in: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (ACL-IJCNLP), 2009, pp. 351–359.
- [27] M. A. Covington, A fundamental algorithm for dependency parsing, in: Proceedings of the 39th Annual ACM Southeast Conference, 2001, pp. 95–102.
- [28] C. Gómez-Rodríguez, Natural language processing and the Now-or-Never bottleneck, Behavioral and Brain Sciences 39 (2016) e74.
doi:10.1017/S0140525X15000795.
URL http://journals.cambridge.org/article_S0140525X15000795

- [29] R. Ferrer-i-Cancho, Euclidean distance between syntactically linked words, *Physical Review E* 70 (2004) 056135.
- [30] R. Rosa, J. Mašek, D. Mareček, M. Popel, D. Zeman, Z. Žabokrtský, HamleDT 2.0: Thirty dependency treebanks stanfordized, in: N. C. C. Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk, S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association (ELRA), Reykjavik, Iceland, 2014.
- [31] J. Nivre, D. Fernández-González, Arc-eager parsing with the tree constraint, *Computational Linguistics* 40 (2) (2014) 259–267.
URL <http://www.aclweb.org/anthology/J/J14/J14-2002.pdf>