ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews ••• (••••) •••–•••

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Comment

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

Received 30 May 2017; accepted 30 May 2017

Communicated by L. Perlovsky

Keywords: Dependency distance; Natural language parsing; Dependency parsing; Crossing dependencies; Non-projectivity

1. Introduction

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 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.

2. 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 [3–5], 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. One 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

DOI of original article: http://dx.doi.org/10.1016/j.plrev.2017.03.002.
E-mail address: cgomezr@udc.es.

http://dx.doi.org/10.1016/j.plrev.2017.05.007

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 [13–15,4,16,5]. 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.

3. 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 [20–24], 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

Please cite this article in press as: Gómez-Rodríguez C. On the relation between dependency distance, crossing dependencies, and parsing. Phys Life Rev (2017), http://dx.doi.org/10.1016/j.plrev.2017.05.007