Are crossing dependencies really scarce?

R. Ferrer-i-Cancho[1,*], C. Gómez-Rodríguez[2], and J. L. Esteban[3]

[1]Complexity & Quantitative Linguistics Lab, LARCA Research Group
Departament de Ciències de la Computació
Universitat Politècnica de Catalunya
Campus Nord, Edifici Omega, Jordi Girona Salgado 1-3
08034 Barcelona, Catalonia (Spain)
[2]Universidade da Coruña
FASTPARSE Lab, LyS Research Group
Departamento de Computación
Facultade de Informática, Elviña
15071 A Coruña, Spain
[3]Logic and Programming, LOGPROG Research Group
Departament de Ciències de la Computació
Universitat Politècnica de Catalunya
Campus Nord, Edifici Omega, Jordi Girona Salgado 1-3
08034 Barcelona, Catalonia (Spain)
[*]Corresponding author, rferrericancho@cs.upc.edu.

October 11, 2017

Abstract

The syntactic structure of a sentence can be modelled as a tree, where vertices correspond to words and edges indicate syntactic dependencies. It has been claimed recurrently that the number of edge crossings in real sentences is small. However, a baseline or null hypothesis has been lacking. Here we quantify the amount of crossings of real sentences and compare it to the predictions of a series of baselines. We conclude that crossings are really scarce in real sentences. Their scarcity is unexpected by the hubiness of the trees. Indeed, real sentences are close to linear trees, where the potential number of crossings is maximized.

Keywords: spatial networks, syntactic dependency trees, crossings, baselines.

1   Introduction

Central to network theory is the definition of null models that shed light on the nature and the significance of network properties [1, 2]. A prototypical example of measure is $\mathcal{T}$, the clustering coefficient of a network (the average proportion of

pairs of neighbours of a vertex that are connected) [3]. It is well known that real networks typically exhibit $\mathcal{T} \gg \mathcal{T}_{ER}$, where $\mathcal{T}_{ER}$ is the clustering coefficient of an Erdős-Rényi graph with the same density of links [3]. In this setup, $\mathcal{T}_{ER} = \delta$ where δ is the density of links of the real network. As real networks are sparse, δ is a small number while $\mathcal{T}$ is typically a large number (a number close to 1). Hence $\mathcal{T}$ is much greater than expected by chance.

As a null hypothesis, the Erdős-Rényi graph involves minimal information from a real network: its number of vertices and its number of links. In an attempt to understand the origins of the properties of real networks, researchers have been defining null hypotheses that are stronger than the Erdős-Rényi graph in the sense that they involve more information from a real network. Perhaps the most popular example are random graphs with a degree sequence that matches that of the real graph. Various models that differ in how they sample the space of possible graphs have been designed. One is the configuration model or pairing model, a model that has been very successful from a theoretical perspective [4, 5, 6] but has very limited applicability as a baseline for real networks due to its inefficiency [2]. The configuration model samples uniformly on the space of configurations (pairings of stubs) [2]. Another example is the switching model, a model that produces a random graph from a given graph preserving the degree sequence as in the configuration model [7, 8, 9]. The switching model can be configured to sample uniformly over the space of possible graphs with the same degree sequence [10, 11].

Here we focus on baselines and null hypotheses for a particular kind of network, i.e., the syntactic structure of sentences, where nodes correspond to words and connections indicate syntactic dependencies between elements, e.g., the dependency between the subject of a sentence and the corresponding verb (Fig. 1) [12]. Syntactic dependency networks are typically trees [13, 14, 12] and constitute a particular case of spatial or geographical network [15, 16, 17] in one dimension, the dimension defined by the linear order of the words in the sentence. The specific measure which we aim to compare against null hypotheses is the number of crossings, which we will denote by C in general. Suppose that vertices are arranged sequentially and that $\pi(v)$ is the position of vertex v in the sequence ($\pi(v) = 1$ for the first vertex of the sequence, $\pi(v) = 2$ for the second vertex of the sequence, and so on). Suppose that we have two edges $u \sim v$ and $s \sim t$ such that $\pi(u) < \pi(v)$ and $\pi(s) < \pi(t)$. We say that $u \sim v$ and $s \sim t$ cross if and only if $\pi(u) < \pi(s) < \pi(v) < \pi(t)$ or $\pi(s) < \pi(u) < \pi(t) < \pi(v)$. With this definition one can count C_{true}, the observed number of crossings for a given sentence. The top of Fig. 1 shows a planar sentence, i.e., a sentence without crossings ($C_{true} = 0$), whereas the bottom shows an ordering of the same sentence with one dependency crossing ($C_{true} = 1$), involving the dependency between "Yesterday" and "arrived" and the dependency between "woman" and "who".

It is well known that crossing dependencies, those that cross each other when drawn above the words of a sentence, are relatively uncommon in natural language [19, 14]. It is widely accepted that the number of crossings of real sentences is small [19, 14, 12, 20, 21, 22, 23]. A challenge for the belief that
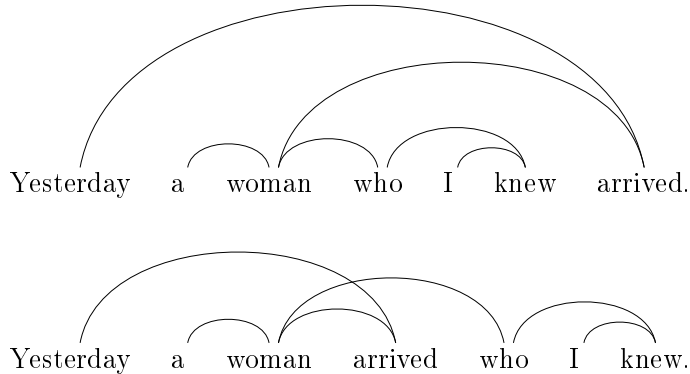
2

Figure 1: Top: The syntactic dependency structure of a sentence without crossings. Bottom: The syntactic dependency structure of the same sentence with an ordering that produces one crossing. Borrowed from [18].

the number of crossings is really small is that the proportion of sentences of a corpus that are not planar, namely, they have at least one crossing, can be very large. For instance, about 30% of sentences in German and Dutch corpora are not planar (Table 1 of [24]). Another challenge is that the scarcity of crossings is not supported with a baseline or null hypothesis. For instance, a star tree (a tree where all connections are formed with a hub vertex as on top of Fig. 2) cannot have crossings [25]. Therefore, reaching the theoretical minimum number of crossings does not suffice to conclude that the number of crossings is smaller than expected by chance: for a star tree, it could not be otherwise. In other words, the number of crossings of a star tree is really small (it is minimum) but not scarce with respect to all possible linear orderings of its vertices.

In this article, we aim to quantify the actual number of crossings of real sentences and to clarify the issue of the presumable scarcity of crossing dependencies. More specifically, we will calculate the actual number of crossings in large collections of sentences and compare them against the predictions of baselines and null hypotheses that vary in the amount of information that they involve about a real tree, as it happens with null hypotheses for real networks.

The remainder of the article is organized as follows. Section 2 presents a series of baselines that will be used to assess if the actual number of crossings in sentences is really scarce. Some baselines are borrowed from previous research [26, 27] while others are introduced here. It also presents a measure of hubiness (a normalized measure of the similarity between a dependency tree and a star tree) and shows the relationship between that measure and the potential number of crossings of a tree. Section 3 presents the collections of dependency networks from different languages that will be used in Section 4 to compare the actual
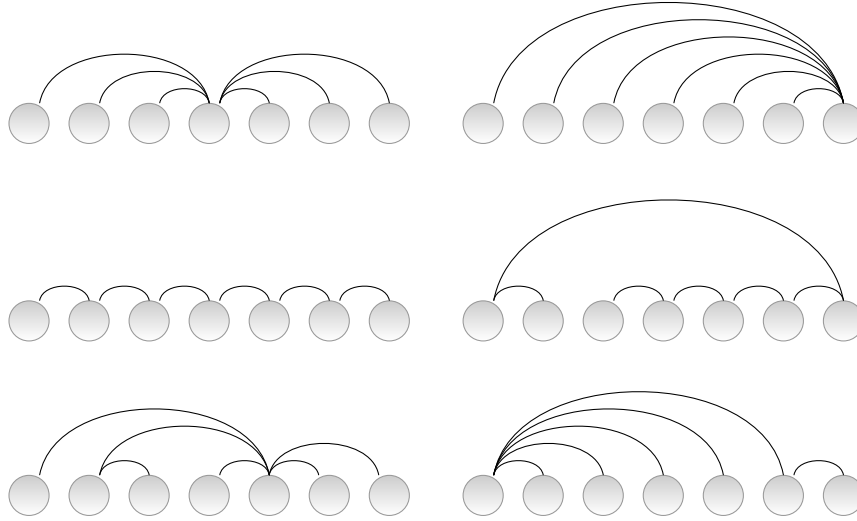
Figure 2: Linear arrangements of trees with $n = 7$ vertices. Top: A star tree. Center: A linear tree. Bottom: A quasi-star tree.

number of crossings of real dependency trees against the random baselines of Section 2. Section 4 also analyzes the degree of hubiness of real dependency trees. Section 5 discusses the results.

2   Baselines for the number of crossings

2.1   Absolute baselines for the number of crossings

Star trees and linear trees are crucial to understand the limits of the variation of C. A star tree is a tree where a vertex has maximum degree (namely $n - 1$, an thus all other vertices have degree 1) [25]. A linear tree is a tree where vertex degrees do not exceed 2 (and therefore all vertices have degree 2 except a couple that have degree 1) [25]. See examples of star and linear trees in Fig. 2.

When looking for a reference for the actual number of crossings of a sentence, a first step is to calculate the potential number of crossings. In a syntactic dependency tree of n nodes, the number of edges is $n - 1$ and therefore the total number of crossings cannot exceed

$$\binom{n-1}{2} = \frac{(n-1)(n-2)}{2}. \tag{1}$$

However, this is a rough estimate as edges that share a vertex cannot cross. Taking account this fact, one may define Q, the size of the set of pair of edges

that may potentially cross. $|Q|$, the cardinality of this set, depends on n and $\langle k^2 \rangle$, the second moment of degree about zero, defined as

$$\langle k^2 \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i^2,$$
(2)

with k_i being the degree of the i-th vertex of the network. In particular, one has [25]

$$|Q| = \frac{n}{2} \left(\langle k^2 \rangle_{star} - \langle k^2 \rangle \right),$$
(3)

where

$$\langle k^2 \rangle_{star} = n - 1$$
(4)

is the value of $\langle k^2 \rangle$ for a star tree of n vertices [25]. Indeed, $|Q|$ reaches extreme values in star trees and linear trees. An overview of the arguments follows (see [27] and the Appendix of [28] for further mathematical details).

The variation of $|Q|$ obeys

$$|Q_{star}| \leq |Q| \leq |Q_{linear}|,$$
(5)

where $|Q_{linear}|$ and $|Q_{star}|$ are the value of $|Q|$ in a linear tree and a star tree, respectively. Obviously, the minimum $|Q|$, namely, $|Q| = 0$ is achieved by a star tree because $\langle k^2 \rangle = \langle k^2 \rangle_{star}$ in that case. The maximum value of $|Q|$ is achieved by a linear tree because that tree yields the minimum value of $\langle k^2 \rangle$, namely

$$\langle k^2 \rangle_{linear} = 4 - 6/n$$
(6)

(when $n \geq 2$). Notice that

$$\langle k^2 \rangle_{star} - \langle k^2 \rangle_{linear} = \frac{(n-2)(n-3)}{n}$$
(7)

Therefore, the maximum value of $|Q|$ is

$$\begin{aligned} |Q_{linear}| &= \frac{n \left(\langle k^2 \rangle_{star} - \langle k^2 \rangle_{linear} \right)}{2} \\ &= \frac{1}{2}(n-2)(n-3) \end{aligned}$$
(8)

for $n \geq 3$ ($|Q_{linear}| = 0$ for $n < 3$).

Obviously, $C = |Q_{star}| = 0$ for any linear arrangement of the vertices of a star tree. We wish to check if there are orderings of the vertices of a linear tree where actually $C = |Q_{linear}|$. Suppose that the vertices of a linear tree with $n \geq 2$ are labelled following a depth-first traversal from one of the leaves. This is equivalent to labelling vertices according to their position in a minimum linear arrangement [29]. Fig. 3 shows linear arrangements of small linear trees with maximum C, namely $C = |Q_{linear}|$. These arrangements can be built by placing all vertices with odd labels in ascending order followed by all vertices
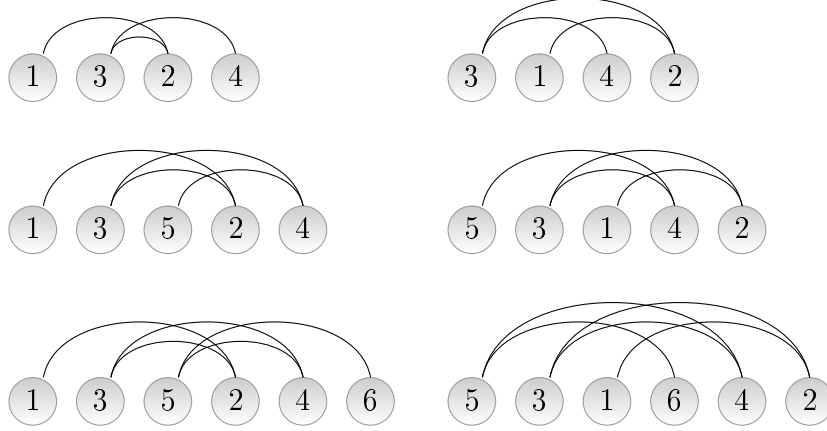
Figure 3: Arrangements of small linear trees with maximum C according to Eq. 8. Top: $n = 4$ and $C = |Q_{linear}| = 1$. Center: $n = 5$ and $C = |Q_{linear}| = 3$. Bottom: $n = 6$ and $C = |Q_{linear}| = 6$.

with even labels in ascending order. By symmetry, it is possible to build other arrangements where $C = |Q_{linear}|$, e.g., placing all vertices with odd labels in descending order followed by all vertices with even labels in descending order. Appendix A presents arrangements that reach $C = |Q_{linear}|$ for a linear tree of an arbitrary size n ($n \geq 3$). As

$$|Q_{linear}| = \binom{n-2}{2},\qquad(9)$$

(recall Eq. 8), it is easy to see how crossing theory replaces the naive upper bound of C in Eq. 1 with a tight one.

Given the results above, the actual number of crossings of a star tree is not surprising at all according to $|Q|$. In contrast, achieving a low number of crossings in a linear tree is unexpected if the tree is sufficiently large.

2.2 A hubiness coefficient

We have seen above that $\langle k^2 \rangle$ is a fundamental structural property of a tree: it determines $|Q|$. As $\langle k^2 \rangle$ determines the range of variation of $|Q|$, $\langle k^2 \rangle$ also determines the solution to the minimum linear arrangement problem: the solution is minimum for linear trees and maximum for star trees [29]. $\langle k^2 \rangle$ is a measure of the hubiness of a tree [25] and we have seen that its range of variation is

$$\langle k^2 \rangle_{linear} \leq \langle k^2 \rangle \leq \langle k^2 \rangle_{star},\qquad(10)$$

where $\langle k^2 \rangle_{linear}$ and $\langle k^2 \rangle_{star}$ are the value of $\langle k^2 \rangle$ in a linear tree and a star tree respectively. The latter allows one to define a hubiness coefficient h as

$$h = \frac{\langle k^2 \rangle - \langle k^2 \rangle_{linear}}{\langle k^2 \rangle_{star} - \langle k^2 \rangle_{linear}} \tag{11}$$

for $n \geq 4$ (for $n < 4$, the only trees that can be formed are both linear and star trees). It is easy to show that $0 \leq h \leq 1$. On the one hand, the fact that

$$\langle k^2 \rangle - \langle k^2 \rangle_{linear} \geq \langle k^2 \rangle_{linear} - \langle k^2 \rangle_{linear}$$
$$= 0 \tag{12}$$

gives

$$h_{linear} = 0 \leq h. \tag{13}$$

On the other hand, the fact that

$$\langle k^2 \rangle - \langle k^2 \rangle_{linear} \leq \langle k^2 \rangle_{star} - \langle k^2 \rangle_{linear} \tag{14}$$

gives

$$h \leq h_{star} = 1. \tag{15}$$

Therefore, h measures the similarity between a tree and a star tree (or the dissimilarity with respect to a linear tree) from the perspective of $\langle k^2 \rangle$. Applying Eqs. 4 and 6 to Eq. 11, one obtains

$$h = \frac{n\left(\langle k^2 \rangle - 4\right) + 6}{n^2 - 5n + 6}$$
$$= \frac{n\left(\langle k^2 \rangle - 4\right) + 6}{(n-2)(n-3)}. \tag{16}$$

Note that h is a normalized degree variance. To see it, recall that the degree variance is

$$V[k] = \langle k^2 \rangle - \langle k \rangle^2 \tag{17}$$

and that $V[k]$ is fully determined by $\langle k^2 \rangle$ because $\langle k \rangle = 2 - 2/n$ [30] for any tree such that $n \geq 1$. Therefore,

$$V[k]_{linear} \leq V[k] \leq V[k]_{star} \tag{18}$$

and

$$h = \frac{V[k] - V[k]_{linear}}{V[k]_{star} - V[k]_{linear}}. \tag{19}$$

h is 1 when the degree variance is maximum and 0 when variance is minimum.

It is also easy to show that h is the complementary of the normalized potential number of crossings, i.e.,

$$h = 1 - \frac{|Q|}{|Q_{linear}|}. \tag{20}$$

7

Therefore, h is 1 when the potential number of crossings is minimum and 0 when it is maximum. Applying the definition of $|Q|$ in Eq. 3 and $|Q_{linear}|$ in Eq. 8 to

$$h = \frac{|Q_{linear}| - |Q|}{|Q_{linear}|}, \tag{21}$$

one recovers Eq. 16 after some algebra.

2.3 Random baselines for the number of crossings

We consider C, the number of crossings of a sentence, in a uniformly random linear arrangement (URLA) of its elements. In this baseline, the expected number of crossings is [26, 27]

$$E_{URLA}[C] = \frac{|Q|}{3}. \tag{22}$$

Another baseline can be obtained assuming that the tree is a uniformly random labelled tree (URLT). This choice improves previous research where random trees that deviate from a uniform distribution were used [20] as a control for C_{true}.

The expected value of $\langle k^2 \rangle$ in a URLT is [26]

$$
\begin{aligned}
E_{URLT}\left[\langle k^2 \rangle\right] &= \left(1 - \frac{1}{n}\right)\left(5 - \frac{6}{n}\right) \\
&= \frac{(n-1)(5n-6)}{n^2}
\end{aligned}
\tag{23}
$$

and then the expected number of crossings in a URLA of an URLT is

$$
\begin{aligned}
E_{URLA}^{URLT}[C] &= E_{URLA}\left[\frac{|Q|}{3}\right] \\
&= \frac{n}{6}\left(n - 1 - E\left[\langle k^2 \rangle\right]\right) \tag{24} \\
&= \frac{(n-1)(n-2)(n-3)}{6n}. \tag{25}
\end{aligned}
$$

Notice that $E_{URLA}^{URLT}[C]$ is related with the unrestricted baseline of the previous subsection. Combining Eqs. 8 and 25, one obtains

$$
\begin{aligned}
E_{URLA}^{URLT}[C] &= \frac{n-1}{3n}|Q_{linear}| \tag{26} \\
&\approx \frac{1}{3}|Q_{linear}| \tag{27}
\end{aligned}
$$

for sufficiently large n. This implies that the expected number of crossings in a random linear arrangement of a URLT is very close to the expected number of crossings in a URLA of a linear tree.

2.4 Random baselines for the hubiness coefficient

Recalling Eqs. 7 and 6, it is easy to see that the expected value of the hubiness coefficient in a uniformly random labelled tree is

$$
\begin{aligned}
E_{URLT}[h] &= E\left[\frac{\langle k^2 \rangle - \langle k^2 \rangle_{linear}}{\langle k^2 \rangle_{star} - \langle k^2 \rangle_{linear}}\right] \\
&= \frac{n\left(E_{URLT}\left[\langle k^2 \rangle\right] - \langle k^2 \rangle_{linear}\right)}{(n-2)(n-3)}.
\end{aligned}
\tag{28}
$$

Recalling 23 and noting that

$$
E_{URLT}\left[\langle k^2 \rangle\right] - \langle k^2 \rangle_{linear} = \frac{(n-2)(n-3)}{n^2},
\tag{29}
$$

we finally obtain

$$
E_{URLT}[h] = \frac{1}{n}.
\tag{30}
$$

The latter implies that, as n tends to infinity, the expected hubiness of URLTs vanishes while the similarity between URLTs and linear trees is maximized. Linear trees swallow practically all probability mass, in agreement with the finding that the expected number of crossings in a URLA of a URLT tends to that of a linear tree as n tends to infinity (Eq. 27). Furthermore, the finding that $E_{URLT}[h] = 1/n$ suggests that the harmonic mean (or its inverse) could be used to evaluate the hubiness of real sentences with respect to URLTs.

2.5 Network theory revisited

In Section 1, we have reviewed various null hypotheses that are used in network theory. The Erdős-Rényi model takes the number of vertices and the number of links of a real network and discards the structure of the real network. The configuration or pairing model and the switching model go a step further incorporating the degree distribution. Our baselines and null hypotheses also parallel this increasing amount of information about the real network that they incorporate.

Recall the two kinds of upper bounds for C_{true} in Section 2. If we consider the structure of the tree under consideration irrelevant (e.g., $\langle k^2 \rangle$), the upper bound is $|Q_{linear}|$ (Eq. 8), the maximum value that $|Q|$ can achieve. This bound parallels the Erdős-Rényi model (notice that in a tree the number of edges is $n-1$ and thus not relevant). If we consider the tree structure relevant, then the upper bound is $|Q|$ (Eq. 3) with $\langle k^2 \rangle$ calculated on the tree under consideration. This bound parallels the configuration or pairing model and the switching model: it involves the degree sequence but knowing $\langle k^2 \rangle$ suffices.

Recall also the two kinds of random baselines for C_{true} in Section 2.3. Neglecting the structure of the tree under consideration (e.g., $\langle k^2 \rangle$), a potential baseline is $E_{URLA}^{URLT}[C]$ (Eq. 25), the expected number of crossings in a uniformly

random tree. This null hypothesis parallels the Erdős-Rényi model. Conditioning on the tree structure, then the potential baseline is $E_{URLA}[C]$ (Eq. 3) with $\langle k^2 \rangle$ taken from the tree under consideration. This null hypothesis parallels the configuration or pairing model and the switching model for involving the degree sequence or a function of it.

Our hubiness coefficient is a normalized $\langle k^2 \rangle$ and we have seen that $\langle k^2 \rangle$ plays a fundamental role in trees: its extremal values determine the limits of the variation of Q and thus also the expected number of crossings in a random linear arrangement of a tree. These values also determine the limits of the variation of D_{min}, the minimum sum of edge lengths in a linear arrangement of a given tree (D_{min} is minimum for linear trees and maximum for star trees) [29]. Such a role is reminiscent of the role played by $\langle k^2 \rangle / \langle k \rangle$ in large complex networks concerning, for instance, the spread of epidemics on a network (e.g. a virus on the Internet): if $\langle k^2 \rangle / \langle k \rangle$ diverges the pandemics cannot be stopped [31]. As $\langle k \rangle = 2 - 2/n$ in trees [30], our work extends the importance of $\langle k^2 \rangle / \langle k \rangle$ to the domain of trees.

3   Materials and methods

We aim to compare C_{true} against the different baselines with the help of dependency treebanks. A dependency treebank is a collection, or corpus, of sentences where a dependency graph is provided for every sentence. Our treebanks come from version 2.0 of the HamleDT collection of treebanks [32, 33]. This collection harmonizes previously existing treebanks for 30 different languages into two widely-used annotation guidelines: Universal Stanford dependencies [34] and Prague dependencies [35]. Therefore, this resource allows us to evaluate the baselines not only across a wide range of languages of different families, but also across two well-known annotation schemes. This is useful because observations like the number of dependency crossings in a sentence not only depend on the language, as they are also influenced by annotation criteria ([28] review some examples of how C can be affected by annotation criteria).

As preprocessing, we removed nodes corresponding to punctuation from the analyses in the treebanks, following common practice in research related to statistical properties of dependency structures (e.g. [36, 37]), which is only concerned with dependency relations between actual words. Null elements, which are present in the Bengali, Hindi and Telugu corpora, were also removed as they do not correspond to words. To preserve the structure of the rest of the tree after removing these nodes, non-deleted nodes that had a deleted node as their head were reattached as dependents of their nearest non-removed ancestor. The size of the tree that is obtained corresponds to the length of the sentence in words.

After this preprocessing, we included in our analyses those syntactic dependency structures that (1) defined a tree with at least 4 nodes, and such that (2) the tree was not a star tree. The reason for (1) is that our baselines assume a tree structure [27, 26] and that we wished to avoid the statistical problem of

mixing trees with other kinds of graphs, e.g., the potential number of crossings depends on the number of edges [25, 38, 27]. We focus on trees of at least 4 nodes because for $n < 4$, the number of crossings is always zero. The reason for (2) is that a star tree cannot have crossing dependencies [25]. Ratios with C in the numerator and $|Q|$ in the denominator, e.g., the relative number of crossings, $C/|Q|$ [26], are not defined because $C = |Q| = 0$. Tables 5 and 6 show $p(star)$, the proportion of trees that are star trees (this proportion is calculated after applying condition (1)). On average, this proportion is smaller than 5%.

As star trees are excluded, the random baselines on uniformly random trees must be adapted (see Appendix B for further details). $E_{URLA}^{URLT}[C]$ is replaced by the same expectation conditioning on the fact that star trees are excluded, i.e.

$$E_{URLA}^{URLT}[C|\neg\text{star}] = \frac{(n-1)(n-2)(n-3)}{6(n-n^{4-n})}. \tag{31}$$

It is easy to see that

$$E_{URLA}^{URLT}[C|\neg\text{star}] \approx E_{URLA}^{URLT}[C] \tag{32}$$

for sufficiently large n (compare Eqs. 25 and 31).

The same applies to $E_{URLT}[h]$, that has to be replaced by

$$E_{URLT}[h|\neg\text{star}] = \frac{n^{n-4}-1}{n^{n-3}-1}. \tag{33}$$

It is easy to see that

$$E_{URLT}[h|\neg\text{star}] \approx E_{URLT}[h] = \frac{1}{n} \tag{34}$$

for sufficiently large n.

The corrected versions of the random baselines are expected to matter especially in treebanks with a sufficient concentration of sentences near $n = 4$.

To assess if the number of crossings in our dataset is significantly small, we conducted two Monte Carlo tests for each treebank, corresponding to each of the two random models of trees. In the first test, we evaluated the significance of the observed values of $\langle C_{true} \rangle$ for each treebank with respect to URLTs, by generating randomized versions of the corpora where each tree is replaced by an URLT with the same number of nodes. To generate each URLT, we produced a uniformly distributed Prüfer code [39] and then converted it to a tree (as an implementation with the Aldous-Broder algorithm [40, 41] proved too slow). The Monte Carlo procedure is used to estimate left-P, the probability that a randomized corpus yields a value of $\langle C_{true} \rangle$ that is at least as small as the original one. One concludes that $\langle C_{true} \rangle$ is significantly small if left-P is small enough. In the second test, we evaluated the significance of $\langle C_{true} \rangle$ with respect to URLAs of the trees in the treebank, by generating randomized versions of the treebanks where each syntactic tree is replaced by an URLA of itself. left-P is estimated as in the 1st test. Each test is based on 10^4 randomizations of the
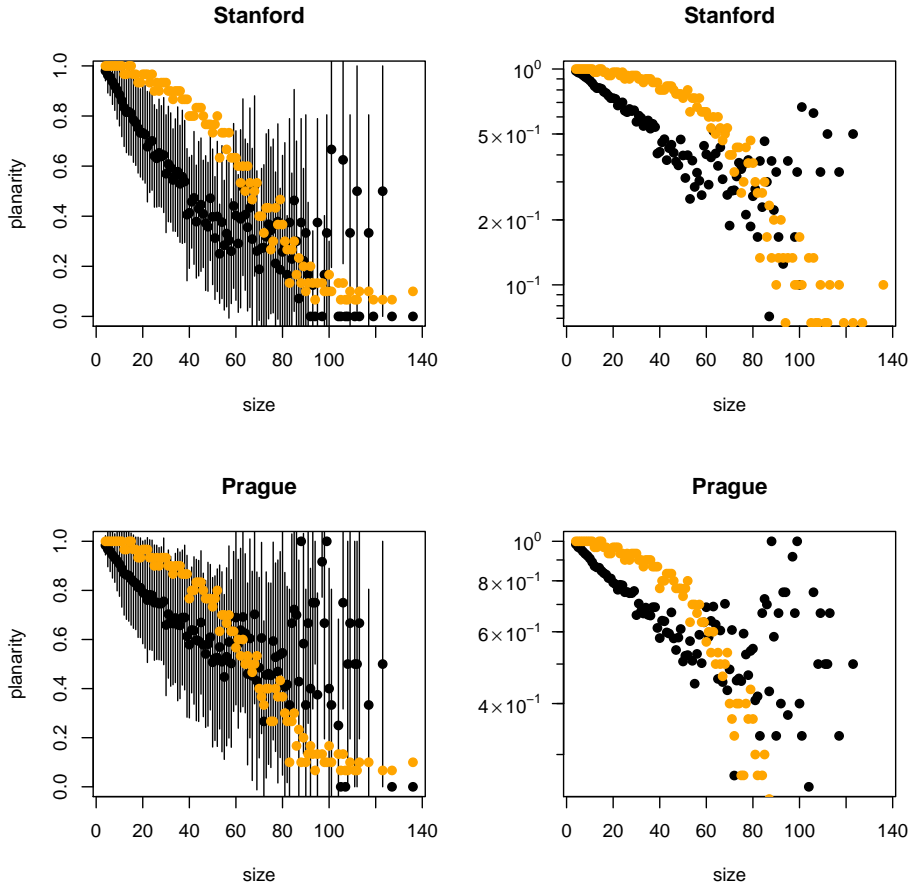
Figure 4: $p(C_{true} = 0)$, the proportion of planar sentences (black), as a function of n, the tree size. The proportion of treebanks having at least one tree of size n is also shown (orange). In the left plots, points and error bars indicate, respectively, mean values and ± 1 standard deviation over proportions in a collection of treebanks. In the right plots, error bars are omitted. Tree sizes represented by less than two treebanks are excluded. Therefore the smallest proportion of treebanks above is 1/15. Top: Stanford annotations. Bottom: Prague annotations.

treebank. Notice that these tests preserve the distribution of tree sizes of the original treebank, that is required to evaluate the significance of a measurement over a whole treebank accurately [42].

We also performed the same couple of tests to evaluate the significance of $p(C_{true} = 0)$, the proportion of planar sentences of a treebank. To evaluate the significance of $\langle h \rangle$, we used URLTs as in the 1st test to estimate left-P and also right-P (the latter being the probability that the randomized corpus yields a value of $\langle C_{true} \rangle$ that is at least as large as the original one).

4 Results

The claim that dependency crossings are scarce in real sentences can be evaluated with at least two statistics. Firstly, $p(C_{true} = 0)$, the proportion of planar sentences (sentences without crossings). $p(C_{true} = 0)$ tends to decrease as n increases on average for all treebanks (Fig. 4). A detailed analysis over all tree sizes shows that this number varies substantially across treebanks (Tables 1 and 2). It is minimum in Ancient Greek with $p(C_{true} = 0) \approx 0.3$ while it reaches its theoretical maximum value ($p(C_{true} = 0) = 1$) for Japanese and Romanian with Prague dependencies. The second smallest proportion of planar sentences is achieved by Latin with $p(C_{true} = 0) \approx 0.5$, followed by German and Dutch with $p(C_{true} = 0)$ slightly below 0.7. Our findings are consistent with a previous report of 30% of sentences in German and Dutch that are not planar (Table 1 of [24]).

Secondly, one can look at the behavior of the actual number of crossings. C_{true} tends to increase as n increases over all treebanks (Fig 5). Interestingly, the plots in double logarithmic scale reveal the presence of a breakpoint at $n = 13$ that separates an initial regime of fast growth of C_{true} from a second regime of slower growth (Fig 5). Hereafter, we will use $\langle ... \rangle$ over a tree measure to indicate a mean over the whole ensemble of sentences of a treebank included in our analysis. Although the proportion of planar sentences can be very low when putting all tree sizes together, the number of crossings is apparently small: $\langle C_{true} \rangle$ does not reach 3.4 in any of the treebanks (Tables 1 and 2). $\langle C_{true} \rangle$ is above 1 in only three languages: Ancient Greek, Latin and Dutch for Stanford dependencies; and only Ancient Greek and Latin for Prague dependencies. These average numbers of observed crossings are really small when compared against the average potential number of crossings of a linear tree of the same size ($\langle |Q_{linear}| \rangle$) or the average potential number of crossings of the same tree ($\langle |Q| \rangle$). In order of magnitude, the difference between $\langle |Q_{linear}| \rangle$ and $\langle Q \rangle$ is small, suggesting that real trees are close to linear trees, namely, their hubiness is low.

A deeper evaluation of the scarcity of crossing dependencies can be made with the help of ratios between C_{true} and the different baselines: $C_{true}/|Q_{linear}|$, $C_{true}/|Q|$, $C_{true}/E_{URLA}[C]$ and $C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}]$. $C_{true}/|Q|$ has already been used in research on crossings in random trees [26]. Bear in mind that

- All these ratios are positive but only $C_{true}/|Q_{linear}|$ and $C_{true}/|Q|$ are

13
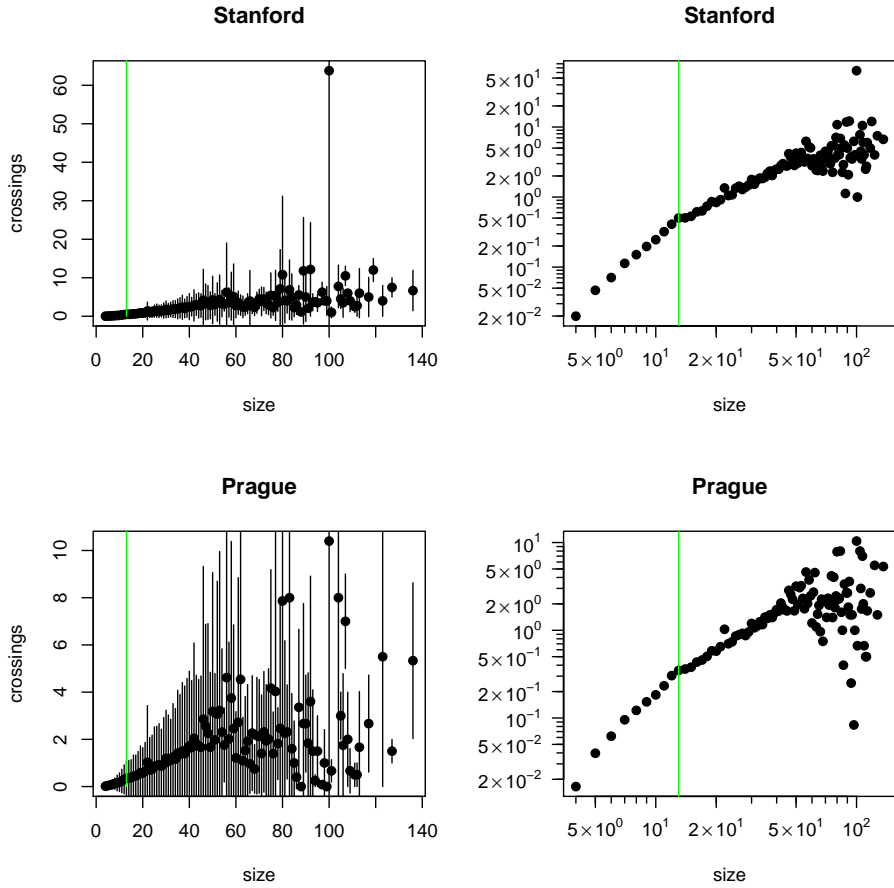
Figure 5: $\langle C_{true} \rangle$, the mean number of dependency crossings (black) as a function of n, the tree size. As a guide to the eye, a vertical line (green) for $n = 13$ is also shown. Top: Stanford annotations. Bottom: Prague annotations. The format is the same as in Fig. 4.

Figure 6: $C_{true}/|Q|$, the relative number of crossings with respect to the potential number of crossings of the tree (black) and $C_{true}/|Q_{linear}|$, the relative number of crossings with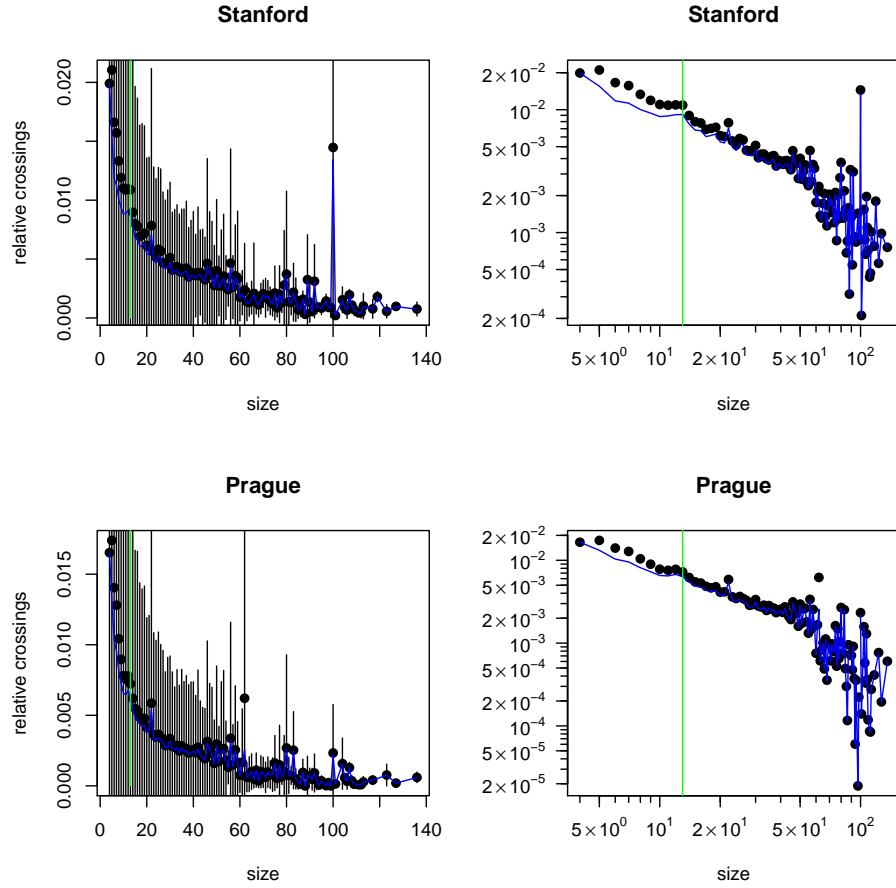 respect to the potential number of crossings of a linear tree (blue), as a function of n, the tree size. As a guide to the eye, a vertical line (green) for $n = 13$ is also shown. The format is the same as in Fig. 4.

15

bounded above by 1.

- Each ratio defined on random baselines is proportional or approximately proportional to a deterministic baseline. On the one hand,

$$\frac{C_{true}}{E_{URLA}[C]} = 3\frac{C_{true}}{|Q|} \tag{35}$$

thanks to Eq. 22. On the other hand,

$$\frac{C_{true}}{E_{URLA}^{URLT}[C|\neg\text{star}]} = \frac{3(n - n^{4-n})}{n-1}\frac{C_{true}}{|Q_{linear}|} \tag{36}$$

thanks to Eqs. 8 and 31. Then

$$\frac{C_{true}}{E_{URLA}^{URLT}[C|\neg\text{star}]} \approx 3\frac{C_{true}}{|Q_{linear}|} \tag{37}$$

for sufficiently large n.

- Although

$$\frac{C_{true}}{|Q_{linear}|} \leq \frac{C_{true}}{|Q|} \tag{38}$$

thanks to $|Q| \leq |Q_{linear}|$, the relationship between

$$\frac{C_{true}}{E_{URLA}[C]} \tag{39}$$

and

$$\frac{C_{true}}{E_{URLA}^{URLT}[C|\neg\text{star}]} \tag{40}$$

is uncertain.

$C_{true}/|Q_{linear}|$ and $C_{true}/|Q|$ tend to decrease as tree size increases (Fig. 6) and the same is expected to happen to their corresponding random baselines thanks to the proportionality relationships above (Eqs. 35 and 37). Therefore, the evidence of the scarcity of crossings increases as tree size increases.

The ratios in Tables 3 and 4 show that, on average, the actual number of crossings is smaller than that of the baseline for all treebanks and for all baselines: all average ratios are below 0.3. These ratios allow one to analyze with more detail the difference in magnitude between C_{true} and the different baselines (Tables 3 and 4):

- $\langle C_{true}/|Q_{linear}|\rangle$ indicates that, on average, C_{true} is at least 10 times smaller than $|Q_{linear}|$ and $|Q|$ across languages. The smallest differences are achieved by Ancient Greek, where $\langle C_{true}/|Q_{linear}|\rangle \approx 0.07$ and $\langle C_{true}/|Q|\rangle \approx 0.09$. The relative number of crossings with respect to the same tree, i.e., $C_{true}/|Q|$, is expected to be about 1/3 in a random linear arrangement of vertices [26] but indeed it is much smaller.

16

- $\langle C_{true}/|Q_{linear}| \rangle \leq \langle C_{true}/|Q| \rangle$ as expected but the difference between $\langle C_{true}/|Q_{linear}| \rangle$ and $\langle C_{true}/|Q| \rangle$ is small, suggesting that real trees are closer to linear trees than to star trees.

- $\langle C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}] \rangle$ indicates that, on average, C_{true} is at least 10 times smaller than $E_{URLA}^{URLT}[C|\neg\text{star}]]$ across treebanks except for Ancient Greek and Latin. For Ancient Greek, $\langle C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}] \rangle \approx 0.23$ on average with Stanford dependencies and $\langle C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}] \rangle \approx 0.24$ with Prague dependencies. For Latin, $\langle C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}] \rangle \approx 0.14$ on average with Stanford dependencies and $\langle C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}] \rangle \approx 0.13$ with Prague dependencies.

- $\langle C_{true}/E_{URLA}[C] \rangle$ indicates that, on average, C_{true} is at least 10 times smaller than $E_{URLA}[C]$ across treebanks except for Ancient Greek and Latin. For Ancient Greek, $\langle C_{true}/E_{URLA}[C] \rangle \approx 0.27$ on average with both Stanford and Prague dependencies. For Latin, $\langle C_{true}/E_{URLA}[C] \rangle \approx 0.15$ with Stanford dependencies and $\langle C_{true}/E_{URLA}[C] \rangle \approx 0.14$ with Prague dependencies.

- The difference between $\langle C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}] \rangle$ and $\langle C_{true}/E_{URLA}[C] \rangle$ is small. The condition $\langle C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}] \rangle \leq \langle C_{true}/E_{URLA}[C] \rangle$ holds for all treebanks with Stanford annotations, as well as for all treebanks with Prague annotations except for Japanese and Persian.

The significance of the gap that separates the actual number of crossings and the predictions of random baselines must be evaluated statistically. Indeed, $p(C_{true} = 0)$ and $\langle C_{true} \rangle$ are smaller than expected by URLTs and URLAs: the Monte Carlo test described in Section 3 yields left-$P < 10^{-4}$ for all the treebanks and both random baselines.

Fig. 7 shows that the hubiness of trees tends to decrease as n increases. Tables 5 and 6 also show that $\langle h \rangle$ never exceeds 0.24 and is ≈ 0.1 across treebanks, suggesting that real trees are closer to linear trees than to star trees. The similarity between linear trees and real trees supports the little difference reported above between $\langle C_{true}/|Q_{linear}| \rangle$ and $\langle C_{true}/Q \rangle$. Indeed, recall the alternative definition of h in Eq. 20. Concerning URLTs, Fig. 7 shows that the average hubiness of real sentences tends to be above the average hubiness that is expected in a URLT over the ensemble of treebanks. A detailed analysis reveals that the average hubiness of real sentences is above the average hubiness that is expected in a URLT for all treebanks with Stanford dependencies (Table 5). However, this does not hold for the Arabic, Japanese and Persian treebank with Prague dependencies (Table 6) but the difference is small. The systematic deviation between $\langle h \rangle$ and $\langle E_{URLT}[h|\neg\text{star}] \rangle$ suggests that the hubiness of real dependency trees cannot be explained by sampling of URLTs, especially for Stanford dependencies. The gap between URLTs and real syntactic dependency trees is smaller for Prague dependencies, as Fig. 7 suggests. Notice that $\langle h \rangle$ is about twice $\langle E_{URLT}[h|\neg\text{star}] \rangle$ with Stanford dependencies whereas $\langle h \rangle$ is about 1.4 times $\langle E_{URLT}[h|\neg\text{star}] \rangle$ with Prague dependencies. The Monte
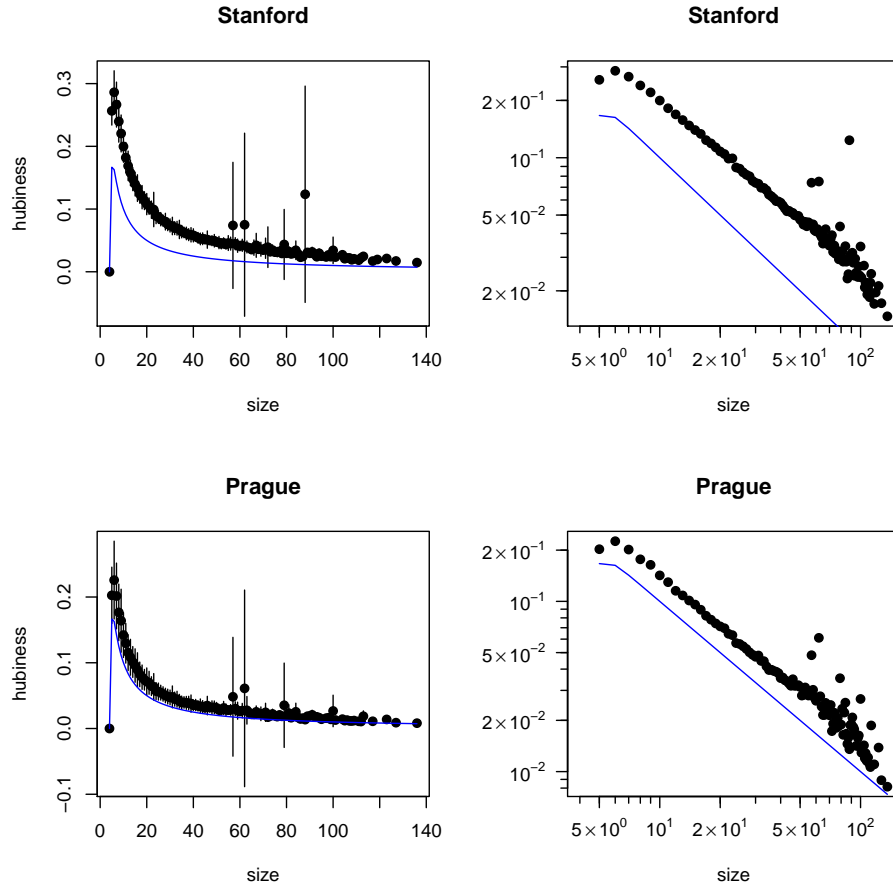
Figure 7: h, the hubiness coefficient (black), and $E_{URLT}[h|\neg\text{star}] \approx 1/n$, the expected hubiness coefficient of a uniformly random labelled tree excluding star trees (blue), as a function of n, the tree size. The format is the same as in Fig. 4. When $n = 4$ a tree can only be a star tree or a linear tree. As our analysis excludes star trees (Section 3), $h = 0$ for $n = 4$. For this reason $n = 4$ is included in normal scale but excluded in log-log scale.

Carlo tests indicate that $\langle h \rangle$ is significantly large in all treebanks with Stanford annotations (right-$P < 10^{-4}$). The results are less homogeneous for Prague annotations: $\langle h \rangle$ is significantly small in Arabic, Japanese and Persian (left-$P < 10^{-4}$) but significantly large for the remainder (right-$P < 10^{-4}$ in all cases except right-$P = 10^{-4}$ for Portuguese).

5 Discussion

We have clarified the issue of the scarcity of crossing dependencies. We have provided the first evidence that the actual number of crossings is significantly small. From the perspective of planarity, the proportion of non-planar sentences can be "high" in certain languages (e.g., Dutch) but still significantly low. On the other hand, the mean number of crossings per sentence is a small number, consistently with the claim that crossings in real sentences are scarce [19, 14, 12, 20, 21, 23] even in languages where non-planar sentences abound. However, whether a number is small or large is a matter of the scale or the units of measurement [43]. Therefore, statistical testing and a theory of crossings (Section 2) are vital. The former shows that crossings are significantly low. The latter helps to understand why and how.

The low number of crossings of real sentences could be trivially explained by a high hubiness, which would immediately lead to a low value of $|Q|$, the potential number of crossings. Fig. 7 indicates that this is unlikely to be the case for sufficiently large trees: the hubiness of trees tends to decrease as n increases and so the relative number of crossings does (Fig. 6). The contribution of hubiness to keeping the number of crossings low decreases as n increases.

Furthermore, the hubiness coefficient never exceeds 25% and is about 10% on average, although it is significantly high with respect to URLTs in the majority of treebanks. The point: is this number large enough to expect a low number of crossings? Thanks to Eq. 20, the relative potential of crossings with respect to a linear tree turns out to be at least 75%, and 90% on average. This strongly suggests that hubiness has a secondary role in explaining the scarcity of crossing dependencies. Indeed, we have seen above that various baselines indicate that real trees are close to linear trees. We have also seen that the gap between real trees and URLTs reduces with Prague annotations. The statistical similarity between real dependency trees and linear trees is what makes the low number of crossing dependencies to be really scarce: linear trees maximize the potential number of crossings, as we have shown above.

The challenge for future research is to determine the true reason for the low number of crossings in sentences. A long standing hypothesis is that the low number of crossings of real sentences is a side effect of the principle of dependency length minimization, namely, the minimization of the distance between linked vertices in the linear sequence [20, 38, 26, 44]. The low hubiness of real sentences suggests that hubiness may have a secondary role in reducing crossing dependencies. We hope that our quantification of the number of crossing dependencies with respect to baselines stimulates further research on the actual

Table 1: A summary of the analysis of the number of crossings in treebanks from different languages, annotated under the Stanford guidelines. First, $p(C_{true} = 0)$, the proportion of planar sentences of the treebanks. Second, mean and standard deviation over the whole treebank for a series of measures: C_{true}, the actual number of crossings, $|Q_{linear}|$, the maximum potential number of crossings and $|Q|$, the potential number of crossings. For every metric (except $p(C_{true} = 0)$), we show $\mu \pm \sigma$, where μ is the average value of the measures over all the trees of the treebank included in our analysis and σ is their standard deviation.

Language	$p(C_{true} = 0)$	C_{true}	$|Q_{linear}|$	$|Q|$
Arabic	0.690	0.981 ± 2.059	505.0 ± 967.9	486.2 ± 949.1
Basque	0.932	0.139 ± 0.684	56.4 ± 63.8	48.3 ± 58.2
Bengali	0.944	0.106 ± 0.627	16.9 ± 24.4	13.4 ± 21.4
Bulgarian	0.843	0.360 ± 1.024	90.1 ± 150.0	80.9 ± 143.3
Catalan	0.790	0.642 ± 1.606	394.0 ± 476.7	366.9 ± 460.1
Czech	0.780	0.528 ± 1.277	135.7 ± 193.3	124.0 ± 184.5
Danish	0.734	0.680 ± 1.482	155.6 ± 223.4	141.6 ± 212.8
Dutch	0.654	1.398 ± 2.690	95.6 ± 148.1	86.0 ± 139.9
English	0.787	0.524 ± 1.370	233.8 ± 243.1	213.7 ± 231.0
Estonian	0.974	0.038 ± 0.267	19.6 ± 41.3	15.9 ± 36.8
Finnish	0.891	0.318 ± 1.215	62.4 ± 79.0	54.6 ± 73.4
German	0.684	0.783 ± 1.593	150.9 ± 229.4	138.2 ± 219.9
Greek (Anc.)	0.312	3.262 ± 4.880	89.1 ± 176.0	77.0 ± 147.1
Greek (Mod.)	0.752	0.654 ± 1.513	289.2 ± 374.2	269.7 ± 361.0
Hindi	0.858	0.304 ± 0.988	202.8 ± 222.5	183.3 ± 211.0
Hungarian	0.728	0.972 ± 2.426	190.8 ± 245.4	173.7 ± 232.5
Italian	0.851	0.415 ± 1.309	199.0 ± 341.3	183.9 ± 329.2
Japanese	0.884	0.164 ± 0.523	51.2 ± 95.3	46.3 ± 90.6
Latin	0.505	2.179 ± 3.920	115.1 ± 179.8	102.7 ± 168.3
Persian	0.785	0.591 ± 2.833	130.8 ± 361.8	121.2 ± 351.9
Portuguese	0.751	0.634 ± 1.452	252.8 ± 401.2	236.2 ± 388.8
Romanian	0.953	0.102 ± 0.474	52.6 ± 96.5	46.8 ± 91.1
Russian	0.810	0.417 ± 1.131	119.8 ± 194.0	109.5 ± 185.9
Slovak	0.819	0.456 ± 1.267	109.9 ± 234.2	99.3 ± 225.4
Slovenian	0.775	0.775 ± 2.068	143.4 ± 257.1	127.6 ± 244.2
Spanish	0.794	0.622 ± 1.594	397.4 ± 432.6	371.2 ± 416.8
Swedish	0.824	0.487 ± 1.764	133.2 ± 209.8	120.7 ± 199.6
Tamil	0.979	0.024 ± 0.174	100.3 ± 141.8	89.6 ± 133.3
Telugu	0.986	0.014 ± 0.117	5.5 ± 7.7	4.3 ± 6.3
Turkish	0.945	0.098 ± 0.524	74.5 ± 134.4	66.9 ± 126.8
Macro avg	0.800	0.622 ± 1.495	152.4 ± 231.5	140.0 ± 221.3

Table 2: A summary of the analysis of the number of crossings in treebanks from different languages, annotated under the Prague guidelines. The format and data shown are as in Table 1.

Language	$p(C_{true} = 0)$	C_{true}	$|Q_{linear}|$	$|Q|$
Arabic	0.945	0.088 ± 0.458	507.6 ± 968.3	496.0 ± 956.1
Basque	0.933	0.125 ± 0.641	55.6 ± 63.3	49.3 ± 59.3
Bengali	0.939	0.124 ± 0.693	15.9 ± 21.4	12.6 ± 18.8
Bulgarian	0.905	0.125 ± 0.426	88.3 ± 147.5	83.0 ± 143.7
Catalan	0.955	0.087 ± 0.532	392.9 ± 476.4	376.4 ± 466.9
Czech	0.785	0.373 ± 0.920	132.2 ± 191.4	124.7 ± 186.0
Danish	0.880	0.164 ± 0.519	154.3 ± 222.2	146.3 ± 216.1
Dutch	0.673	0.990 ± 1.928	94.1 ± 147.4	88.6 ± 142.9
English	0.941	0.107 ± 0.811	233.2 ± 248.3	220.4 ± 241.2
Estonian	0.992	0.013 ± 0.157	19.0 ± 41.0	16.0 ± 37.9
Finnish	0.908	0.128 ± 0.476	61.6 ± 78.8	56.1 ± 75.2
German	0.671	0.723 ± 1.489	148.3 ± 227.9	140.0 ± 221.8
Greek (Anc.)	0.323	3.353 ± 4.446	89.5 ± 181.7	79.1 ± 151.6
Greek (Mod.)	0.867	0.206 ± 0.660	286.6 ± 368.4	274.2 ± 360.1
Hindi	0.769	0.387 ± 0.958	201.8 ± 221.7	189.4 ± 215.0
Hungarian	0.738	0.867 ± 2.143	185.9 ± 247.8	172.3 ± 237.6
Italian	0.959	0.062 ± 0.364	196.8 ± 343.3	187.5 ± 335.7
Japanese	1.000	0.000 ± 0.014	49.4 ± 93.9	46.8 ± 91.3
Latin	0.499	1.850 ± 3.171	114.8 ± 180.2	106.2 ± 172.5
Persian	0.817	0.402 ± 2.122	125.3 ± 355.5	120.2 ± 349.8
Portuguese	0.860	0.247 ± 0.780	250.6 ± 400.9	241.6 ± 394.1
Romanian	1.000	0.000 ± 0.000	51.8 ± 95.9	48.0 ± 92.5
Russian	0.907	0.157 ± 0.585	118.9 ± 193.7	112.6 ± 188.8
Slovak	0.853	0.269 ± 0.829	102.9 ± 197.6	96.3 ± 191.3
Slovenian	0.822	0.312 ± 0.802	128.1 ± 224.0	118.4 ± 216.4
Spanish	0.945	0.110 ± 0.625	395.4 ± 432.7	380.2 ± 423.8
Swedish	0.935	0.195 ± 1.476	129.1 ± 202.8	120.0 ± 195.6
Tamil	0.988	0.014 ± 0.130	100.1 ± 141.8	92.2 ± 135.8
Telugu	0.992	0.008 ± 0.089	5.3 ± 7.5	4.1 ± 6.2
Turkish	0.914	0.137 ± 0.540	67.1 ± 123.4	63.2 ± 119.6
Macro avg	0.857	0.388 ± 0.959	150.1 ± 228.2	142.0 ± 221.5

Table 3: A summary of the normalized number of crossings in treebanks from different languages (Stanford annotation): $C_{true}/|Q_{linear}|$, $C_{true}/|Q|$, $C_{true}/E_{URLA}^{URLT}[C|\neg\text{star}]$, $C_{true}/E_{URLA}[C]$. The format is as in Table 1.

Language	$\dfrac{C_{true}}{|Q_{linear}|}$	$\dfrac{C_{true}}{|Q|}$	$\dfrac{C_{true}}{E_{URLA}^{URLT}[C|\neg\text{star}]}$	$\dfrac{C_{true}}{E_{URLA}[C]}$
Arabic	0.00464 ± 0.01769	0.00534 ± 0.02118	0.01523 ± 0.06006	0.01602 ± 0.06353
Basque	0.00300 ± 0.02141	0.00357 ± 0.02519	0.01006 ± 0.07237	0.01070 ± 0.07558
Bengali	0.01023 ± 0.07418	0.01237 ± 0.08446	0.03350 ± 0.23144	0.03711 ± 0.25339
Bulgarian	0.00566 ± 0.02778	0.00685 ± 0.03605	0.01883 ± 0.09411	0.02054 ± 0.10816
Catalan	0.00231 ± 0.01062	0.00267 ± 0.01463	0.00739 ± 0.03577	0.00800 ± 0.04388
Czech	0.00644 ± 0.03073	0.00767 ± 0.03880	0.02123 ± 0.10251	0.02301 ± 0.11639
Danish	0.00731 ± 0.03158	0.00924 ± 0.04160	0.02398 ± 0.10346	0.02771 ± 0.12481
Dutch	0.01867 ± 0.05240	0.02274 ± 0.07002	0.06225 ± 0.18013	0.06821 ± 0.21005
English	0.00263 ± 0.00983	0.00301 ± 0.01226	0.00839 ± 0.03256	0.00904 ± 0.03677
Estonian	0.00213 ± 0.01835	0.00276 ± 0.02492	0.00743 ± 0.06480	0.00827 ± 0.07476
Finnish	0.00615 ± 0.02993	0.00756 ± 0.04028	0.02056 ± 0.10358	0.02269 ± 0.12084
German	0.00745 ± 0.03075	0.00870 ± 0.03777	0.02440 ± 0.10243	0.02611 ± 0.11331
Greek (Anc.)	0.06789 ± 0.11445	0.08954 ± 0.15792	0.22852 ± 0.38975	0.26861 ± 0.47376
Greek (Mod.)	0.00303 ± 0.00991	0.00343 ± 0.01171	0.00969 ± 0.03249	0.01029 ± 0.03512
Hindi	0.00148 ± 0.00650	0.00174 ± 0.00943	0.00472 ± 0.02186	0.00522 ± 0.02828
Hungarian	0.00651 ± 0.02523	0.00768 ± 0.03247	0.02107 ± 0.08304	0.02305 ± 0.09741
Italian	0.00354 ± 0.02341	0.00429 ± 0.03274	0.01172 ± 0.08253	0.01288 ± 0.09822
Japanese	0.00674 ± 0.05244	0.00760 ± 0.05628	0.02213 ± 0.16395	0.02279 ± 0.16883
Latin	0.04057 ± 0.09173	0.05094 ± 0.12026	0.13626 ± 0.31224	0.15281 ± 0.36077
Persian	0.00639 ± 0.03244	0.00765 ± 0.04100	0.02116 ± 0.10850	0.02294 ± 0.12300
Portuguese	0.00438 ± 0.02073	0.00506 ± 0.02530	0.01433 ± 0.06949	0.01519 ± 0.07589
Romanian	0.00182 ± 0.01237	0.00208 ± 0.01443	0.00606 ± 0.04228	0.00625 ± 0.04330
Russian	0.00657 ± 0.03949	0.00782 ± 0.04703	0.02165 ± 0.12800	0.02347 ± 0.14108
Slovak	0.00787 ± 0.04410	0.00959 ± 0.05430	0.02617 ± 0.14497	0.02878 ± 0.16291
Slovenian	0.00901 ± 0.04283	0.01112 ± 0.05502	0.02984 ± 0.14335	0.03335 ± 0.16507
Spanish	0.00233 ± 0.01151	0.00266 ± 0.01462	0.00745 ± 0.03934	0.00797 ± 0.04384
Swedish	0.00485 ± 0.02292	0.00569 ± 0.02809	0.01591 ± 0.07725	0.01708 ± 0.08427
Tamil	0.00044 ± 0.00365	0.00053 ± 0.00437	0.00145 ± 0.01211	0.00158 ± 0.01310
Telugu	0.01049 ± 0.09766	0.01088 ± 0.09861	0.03216 ± 0.29508	0.03263 ± 0.29583
Turkish	0.00298 ± 0.03280	0.00363 ± 0.03662	0.00987 ± 0.10351	0.01090 ± 0.10985
Macro avg	0.00878 ± 0.03465	0.01081 ± 0.04291	0.02911 ± 0.11443	0.03244 ± 0.12873

Table 4: A summary of the normalized number of crossings in treebanks from different languages (Prague annotation). The format and data shown are as in Table 3.

Language	$\frac{C_{true}}{|Q_{linear}|}$	$\frac{C_{true}}{|Q|}$	$\frac{C_{true}}{E_{URLA}^{URLT}[C|\neg star]}$	$\frac{C_{true}}{E_{URLA}[C]}$
Arabic	0.00024 ± 0.00317	0.00026 ± 0.00362	0.00077 ± 0.01022	0.00078 ± 0.01085
Basque	0.00282 ± 0.02033	0.00328 ± 0.02366	0.00945 ± 0.06834	0.00984 ± 0.07097
Bengali	0.01024 ± 0.07027	0.01291 ± 0.08603	0.03417 ± 0.22580	0.03873 ± 0.25808
Bulgarian	0.00364 ± 0.03441	0.00407 ± 0.03688	0.01208 ± 0.10910	0.01221 ± 0.11063
Catalan	0.00024 ± 0.00226	0.00027 ± 0.00269	0.00077 ± 0.00743	0.00081 ± 0.00806
Czech	0.00536 ± 0.02856	0.00605 ± 0.03273	0.01774 ± 0.09367	0.01816 ± 0.09818
Danish	0.00164 ± 0.00817	0.00192 ± 0.01094	0.00538 ± 0.02749	0.00575 ± 0.03281
Dutch	0.01585 ± 0.05088	0.01858 ± 0.06694	0.05325 ± 0.17535	0.05573 ± 0.20081
English	0.00060 ± 0.00594	0.00066 ± 0.00697	0.00194 ± 0.01991	0.00199 ± 0.02090
Estonian	0.00077 ± 0.01239	0.00114 ± 0.01914	0.00268 ± 0.04353	0.00343 ± 0.05742
Finnish	0.00254 ± 0.01391	0.00300 ± 0.01759	0.00850 ± 0.04832	0.00901 ± 0.05277
German	0.00659 ± 0.02447	0.00738 ± 0.02824	0.02157 ± 0.08079	0.02213 ± 0.08473
Greek (Anc.)	0.07030 ± 0.11966	0.09050 ± 0.16363	0.23676 ± 0.40801	0.27151 ± 0.49088
Greek (Mod.)	0.00103 ± 0.00515	0.00112 ± 0.00574	0.00331 ± 0.01716	0.00337 ± 0.01723
Hindi	0.00261 ± 0.00915	0.00295 ± 0.01148	0.00846 ± 0.03122	0.00885 ± 0.03443
Hungarian	0.00597 ± 0.02448	0.00691 ± 0.02964	0.01930 ± 0.07972	0.02075 ± 0.08891
Italian	0.00046 ± 0.00461	0.00053 ± 0.00587	0.00149 ± 0.01538	0.00159 ± 0.01762
Japanese	$7.45 \cdot 10^{-6} \pm 0.00052$	$7.73 \cdot 10^{-6} \pm 0.00053$	$2.48 \cdot 10^{-5} \pm 0.00172$	$2.32 \cdot 10^{-5} \pm 0.00160$
Latin	0.03938 ± 0.09014	0.04750 ± 0.11558	0.13298 ± 0.30915	0.14250 ± 0.34675
Persian	0.00436 ± 0.02215	0.00469 ± 0.02414	0.01444 ± 0.07363	0.01408 ± 0.07242
Portuguese	0.00181 ± 0.01537	0.00200 ± 0.01733	0.00592 ± 0.05020	0.00601 ± 0.05199
Romanian	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
Russian	0.00308 ± 0.02864	0.00349 ± 0.03244	0.01017 ± 0.09206	0.01046 ± 0.09731
Slovak	0.00632 ± 0.04273	0.00740 ± 0.05023	0.02111 ± 0.13945	0.02220 ± 0.15069
Slovenian	0.00531 ± 0.03492	0.00639 ± 0.04142	0.01771 ± 0.11419	0.01918 ± 0.12426
Spanish	0.00039 ± 0.00376	0.00049 ± 0.00632	0.00125 ± 0.01222	0.00146 ± 0.01896
Swedish	0.00146 ± 0.01263	0.00168 ± 0.01503	0.00477 ± 0.04341	0.00504 ± 0.04510
Tamil	0.00036 ± 0.00351	0.00042 ± 0.00412	0.00118 ± 0.01168	0.00125 ± 0.01237
Telugu	0.00804 ± 0.08932	0.00804 ± 0.08932	0.02413 ± 0.26796	0.02413 ± 0.26796
Turkish	0.00571 ± 0.04363	0.00640 ± 0.04716	0.01923 ± 0.14104	0.01921 ± 0.14148
Macro avg	0.00714 ± 0.02750	0.00862 ± 0.03318	0.02381 ± 0.09061	0.02587 ± 0.09954

Table 5: A summary of the hubiness in treebanks from different languages (Stanford annotation): n, the size of the tree, $\langle k^2 \rangle$, the 2nd moment of degree about zero, h, the hubiness coefficient and $E_{URLT}[h|\neg\text{star}]$, the expected h in a uniformly random labelled tree excluding star trees, and $p(star)$ the proportion of trees of the treebank that are star trees. The format is as in Table 1.

Language	n	h	$E_{URLT}[h|\neg\text{star}]$	$p(star)$
Arabic	26.3 ± 21.0	0.089 ± 0.076	0.064 ± 0.045	0.005
Basque	11.8 ± 5.2	0.191 ± 0.120	0.095 ± 0.040	0.044
Bengali	7.5 ± 3.0	0.239 ± 0.160	0.121 ± 0.049	0.163
Bulgarian	13.6 ± 7.6	0.167 ± 0.114	0.089 ± 0.044	0.033
Catalan	26.9 ± 13.8	0.113 ± 0.082	0.049 ± 0.030	0.009
Czech	16.5 ± 8.8	0.135 ± 0.092	0.075 ± 0.040	0.018
Danish	17.2 ± 9.8	0.155 ± 0.109	0.074 ± 0.040	0.032
Dutch	13.6 ± 8.3	0.175 ± 0.124	0.090 ± 0.046	0.033
English	21.7 ± 9.9	0.127 ± 0.084	0.057 ± 0.031	0.011
Estonian	7.3 ± 4.0	0.231 ± 0.169	0.111 ± 0.061	0.237
Finnish	12.2 ± 5.6	0.173 ± 0.113	0.093 ± 0.039	0.026
German	17.2 ± 9.3	0.131 ± 0.089	0.072 ± 0.039	0.034
Greek (Anc.)	13.3 ± 7.8	0.216 ± 0.138	0.093 ± 0.044	0.043
Greek (Mod.)	23.0 ± 12.6	0.116 ± 0.081	0.057 ± 0.034	0.008
Hindi	20.4 ± 9.2	0.140 ± 0.083	0.060 ± 0.029	0.004
Hungarian	19.0 ± 10.4	0.141 ± 0.098	0.068 ± 0.038	0.013
Italian	18.6 ± 11.8	0.144 ± 0.106	0.071 ± 0.042	0.025
Japanese	10.5 ± 6.2	0.155 ± 0.121	0.102 ± 0.050	0.072
Latin	14.9 ± 8.8	0.169 ± 0.113	0.083 ± 0.044	0.028
Persian	14.5 ± 10.9	0.148 ± 0.103	0.088 ± 0.048	0.026
Portuguese	20.9 ± 12.9	0.122 ± 0.091	0.065 ± 0.041	0.019
Romanian	10.7 ± 6.2	0.181 ± 0.124	0.102 ± 0.048	0.056
Russian	15.4 ± 8.6	0.140 ± 0.103	0.080 ± 0.041	0.025
Slovak	14.3 ± 9.0	0.173 ± 0.117	0.087 ± 0.045	0.044
Slovenian	16.1 ± 10.1	0.200 ± 0.126	0.082 ± 0.044	0.052
Spanish	27.1 ± 13.8	0.111 ± 0.085	0.050 ± 0.033	0.012
Swedish	16.2 ± 8.9	0.145 ± 0.090	0.076 ± 0.039	0.013
Tamil	14.6 ± 7.4	0.159 ± 0.103	0.084 ± 0.037	0.008
Telugu	5.4 ± 1.6	0.161 ± 0.166	0.100 ± 0.075	0.305
Turkish	11.8 ± 7.9	0.177 ± 0.130	0.097 ± 0.052	0.078
Macro avg	16.0 ± 9.0	0.157 ± 0.110	0.081 ± 0.043	0.049

Table 6: A summary of the hubiness in treebanks from different languages (Prague annotation). The format and data shown are as in Table 5.

Language	n	h	$E_{URLT}[h|\neg\text{star}]$	$p(star)$
Arabic	26.4 ± 21.0	0.050 ± 0.055	0.064 ± 0.045	0.002
Basque	11.7 ± 5.2	0.163 ± 0.115	0.095 ± 0.041	0.043
Bengali	7.4 ± 2.8	0.232 ± 0.160	0.121 ± 0.049	0.168
Bulgarian	13.4 ± 7.6	0.104 ± 0.101	0.089 ± 0.045	0.021
Catalan	26.9 ± 13.9	0.073 ± 0.067	0.049 ± 0.031	0.006
Czech	16.2 ± 8.7	0.092 ± 0.081	0.075 ± 0.040	0.013
Danish	17.1 ± 9.8	0.096 ± 0.089	0.074 ± 0.041	0.018
Dutch	13.5 ± 8.3	0.110 ± 0.103	0.090 ± 0.047	0.019
English	21.7 ± 10.0	0.086 ± 0.069	0.058 ± 0.032	0.007
Estonian	7.2 ± 4.0	0.211 ± 0.173	0.111 ± 0.062	0.236
Finnish	12.1 ± 5.6	0.133 ± 0.104	0.094 ± 0.040	0.021
German	17.0 ± 9.3	0.091 ± 0.076	0.072 ± 0.039	0.030
Greek (Anc.)	13.3 ± 7.9	0.193 ± 0.141	0.093 ± 0.044	0.040
Greek (Mod.)	22.9 ± 12.5	0.077 ± 0.065	0.057 ± 0.035	0.006
Hindi	20.4 ± 9.2	0.095 ± 0.077	0.060 ± 0.029	0.002
Hungarian	18.7 ± 10.4	0.123 ± 0.097	0.070 ± 0.039	0.015
Italian	18.4 ± 11.8	0.094 ± 0.088	0.071 ± 0.042	0.015
Japanese	10.3 ± 6.2	0.084 ± 0.102	0.100 ± 0.052	0.037
Latin	14.9 ± 8.8	0.124 ± 0.103	0.083 ± 0.044	0.024
Persian	14.2 ± 10.7	0.078 ± 0.084	0.088 ± 0.049	0.012
Portuguese	20.8 ± 12.9	0.067 ± 0.067	0.065 ± 0.041	0.009
Romanian	10.6 ± 6.2	0.128 ± 0.118	0.101 ± 0.048	0.042
Russian	15.3 ± 8.6	0.092 ± 0.086	0.080 ± 0.042	0.018
Slovak	13.9 ± 8.7	0.121 ± 0.108	0.088 ± 0.045	0.037
Slovenian	15.4 ± 9.5	0.150 ± 0.122	0.084 ± 0.044	0.041
Spanish	27.0 ± 13.8	0.069 ± 0.066	0.050 ± 0.033	0.008
Swedish	16.0 ± 8.8	0.112 ± 0.085	0.077 ± 0.039	0.011
Tamil	14.6 ± 7.4	0.134 ± 0.101	0.084 ± 0.037	0.007
Telugu	5.4 ± 1.6	0.156 ± 0.167	0.103 ± 0.075	0.342
Turkish	11.3 ± 7.6	0.111 ± 0.117	0.097 ± 0.054	0.056
Macro avg	15.8 ± 9.0	0.115 ± 0.099	0.081 ± 0.044	0.043

origin of their scarcity and the weight of different factors.

We have observed a breakpoint in the decay of the average number of crossings across treebanks at $n = 13$ (Fig. 5) that is also suggested by the decay of the average relative number of crossings (Fig. 6). We suspect that it could be related to increasing pressure for dependency length minimization for longer sentences. However, the real nature of the breakpoint should be investigated further.

Although the conclusion that crossings in sentences are really scarce does not depend on the annotation format, our analyses indicate that Stanford and Prague dependencies are not statistically equivalent. For instance, we have seen that real trees are closer to URLTs with respect to hubiness when Prague dependencies are considered. This is in line with recent results highlighting various other relevant quantifiable differences between annotation criteria, e.g. in their suitability for automatic parsing [45, 46] or in the prevalence of certain patterns of crossing dependencies [47]. Thus, considering more than one annotation format is useful to analyze underlying properties of syntax, and distinguish them from properties of a specific annotation.

It is worth bearing in mind that syntactic annotation schemes are typically designed based on linguistic considerations [48], as well as technical considerations to facilitate the work of parsers and other language processing systems [49], independently from statistical considerations [50]. Our findings suggest that statistical implications should be involved when improving current annotation formats or developing new ones. Identifying the most appropriate statistical ensemble for syntactic dependency trees is an important problem that should be the subject of future research.

6 Conclusion

We have shown that the number of crossings of real sentences is really scarce with the help of different baselines. Although that scarcity could be easily explained by a high hubiness, the hubiness of real sentences is rather low suggesting that it has a secondary role in the low number of crossings of real sentences. Statistically, syntactic dependency trees seem to be closer to linear trees than to star trees. Our findings provide support for the hypothesis that dependency length minimization is the main force responsible for the scarcity of crossing dependencies.

A The maximum number of crossings of a linear tree

Figure 8 shows arrangements with maximum number of crossings for a series of linear trees of n nodes, with $3 \leq n \leq 7$. Each tree of n nodes is obtained by adding the vertex n to the tree of $n - 1$ nodes. In all cases, the linear ordering of the vertices consists of the odd vertex labels in increasing order, followed by
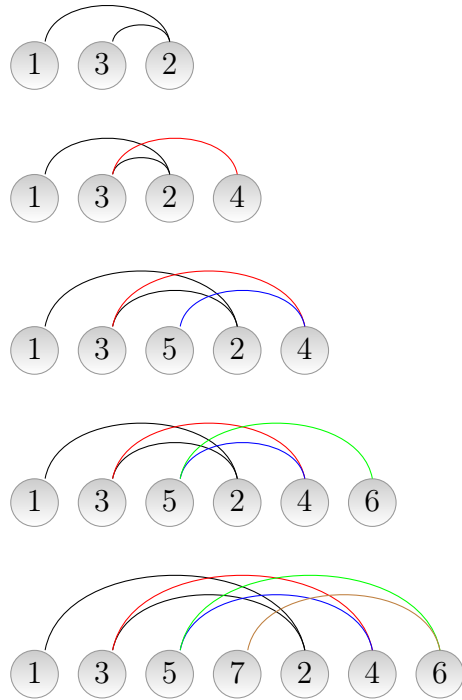
Figure 8: Arrangements of linear trees that maximize the number of crossings. Top to bottom: linear trees of 3, 4, 5, 6 and 7 nodes that have 0, 1, 3, 6 and 10 crossings, respectively. It is easy to check that $C = |Q_{linear}|$ in all cases (recall Eq. 8).

the even vertex labels also in increasing order. We will show that this kind of arrangements achieves the maximum possible number of crossings for linear trees of n nodes. Formally, these orderings can be defined as the sequence of vertices

$$1, 3, \cdots, n + n \bmod 2 - 1, 2, 4, \cdots, n - n \bmod 2. \tag{41}$$

Let $C(n)$ be the corresponding number of crossings. Notice that $C(n) = 0$ for $0 \leq n \leq 3$ [27].

In Figure 8, we adopt the convention that the edge $3 \sim 4$ is always red, $4 \sim 5$ is always blue, $5 \sim 6$ is always green and $6 \sim 7$ is always brown for all linear trees. Thus, it is easy to check the contribution to $C(n)$ of the edge $(n-1) \sim (n)$ with respect to $C(n-1)$; when $n = 4$, the edge $3 \sim 4$ adds one crossing; when $n = 5$, the edge $4 \sim 5$ adds two crossings; when $n = 6$, the edge $5 \sim 6$ adds three crossings and, when $n = 7$, the edge $6 \sim 7$ adds four crossings. After this introduction now comes the proof.

We aim to show that $C(n) = |Q_{linear}|$ (Eq. 8) for $n \geq 3$. First, $C(3) = |Q_{linear}| = 0$, setting the base case. Second, we aim to show that $C(n) = \Delta(n) + C(n-1)$ with $\Delta(n) = n - 3$ for $n \geq 4$. Suppose that a tree of $n-1$ vertices becomes a tree of n vertices adding vertex n and the edge $(n-1) \sim n$. If n is odd, the edge $(n-1) \sim n$ crosses any two edges formed with node i, namely edges $(i-1) \sim i$ and $i \sim (i+1)$, for i even and $2 \leq i \leq n-3$. This yields $\Delta(n) = n - 3$. Note that $(n-1) \sim n$ cannot cross $(n-2) \sim (n-1)$ as they share vertex $n-1$. If n is even, then $(n-1) \sim n$ crosses $1 \sim 2$ and any two edges formed with node i such that i is odd and $3 \leq i \leq n-3$, giving again $\Delta(n) = n - 3$. Therefore,

$$
\begin{aligned}
C(n) &= \sum_{i=4}^{n} \Delta(i) \\
&= \sum_{i=1}^{n-3} i \\
&= \frac{1}{2}(n-2)(n-3)
\end{aligned}
\tag{42}
$$

and finally $C(n) = |Q_{linear}|$ (Eq. 8), as we wanted to prove.

B Expectations on uniformly random labelled trees excluding star trees

There are n labelled star trees: each can be constructed by choosing one of the n vertices as the hub. Since there are n^{n-2} labelled trees in total [51], the probability that a URLT is a star tree is

$$
\begin{aligned}
p(\text{star}) &= \frac{n}{n^{n-2}} \\
&= n^{3-n}.
\end{aligned}
\tag{43}
$$

We define the sum of squared degrees of a tree as [25]

$$K_2 = n \left\langle k^2 \right\rangle \tag{44}$$

and define $p(K_2|\neg\text{star})$ as the probability that a URLT has K_2 as sum of squared degrees knowing that it not a star tree. We have that

$$p(K_2|\neg\text{star}) = \frac{p(\neg\text{star}|K_2)p(K_2)}{p(\neg\text{star})}. \tag{45}$$

We have seen above that the maximum value of $\left\langle k^2 \right\rangle$ for a given n is achieved by a star tree (Eq. 10), and hence the same can be said about the maximum value of K_2. If we call this value K_2^{star}, then

$$p(\neg\text{star}|K_2) = \begin{cases} 1 \text{ for } K_2 < K_2^{star} \\ 0 \text{ for } K_2 = K_2^{star}. \end{cases} \tag{46}$$

Therefore, for $K_2 < K_2^{star}$, we can apply $p(\neg\text{star}) = 1 - p(\text{star})$ and $p(\neg\text{star}|K_2) = 1$ to obtain

$$p(K_2|\neg\text{star}) = \frac{p(K_2)}{1 - p(\text{star})}. \tag{47}$$

If star trees are excluded, the maximum hubiness is reached by a quasi-star tree, a tree that gives the second largest value of $\left\langle k^2 \right\rangle$, and is defined by one vertex of degree $n - 2$, one vertex of degree 2 and the remainder of vertices of degree 1 [38] (Fig. 2). Suppose that K_2^{linear} and $K_2^{quasi-star}$ are the values of K_2 of a linear tree and a quasi-star tree, respectively. The expectation of $\left\langle k^2 \right\rangle$ of a URLT knowing that it is not a star tree is

$$
\begin{aligned}
E_{URLT}\left[\left\langle k^2 \right\rangle|\neg\text{star}\right] &= \frac{1}{n}E_{URLT}\left[K_2|\neg\text{star}\right] \\
&= \frac{1}{n}\sum_{K_2=K_2^{linear}}^{K_2^{quasi-star}} p(K_2|\neg\text{star})K_2 \\
&= \frac{1}{n(1-p(\text{star}))}\sum_{K_2=K_2^{linear}}^{K_2^{quasi-star}} p(K_2)K_2 \\
&= \frac{1}{n(1-p(\text{star}))}\left(E_{URLT}[K_2] - p(\text{star})K_2^{star}\right). (48)
\end{aligned}
$$

Knowing that

$$K_2^{star} = n \left\langle k^2 \right\rangle_{star} = n(n-1), \tag{49}$$

$$
\begin{aligned}
E_{URLT}[K_2] &= nE_{URLT}[\left\langle k^2 \right\rangle] \\
&= \frac{n(n-1)(5n-6)}{n^2} \tag{50}
\end{aligned}
$$

29

thanks to Eq. 23, and recalling Eq. 43, one obtains

$$E_{URLT}\left[\langle k^2 \rangle \middle| \neg \text{star}\right] = \frac{n-1}{1-n^{3-n}}\left(\frac{5n-6}{n^2} - n^{3-n}\right). \tag{51}$$

Notice that

$$E_{URLT}\left[\langle k^2 \rangle \middle| \neg \text{star}\right] = E_{URLT}\left[\langle k^2 \rangle\right] \tag{52}$$

for sufficiently large n (compare Eqs. 23 and 51).

Adapting Eq. 24 to $E_{URLA}^{URLT}[C|\neg \text{star}]$, one obtains

$$E_{URLA}^{URLT}[C|\neg \text{star}] = \frac{n}{6}\left(n-1-E\left[\langle k^2 \rangle \middle| \neg \text{star}\right]\right). \tag{53}$$

Plugging Eq. 51 to 53, one obtains

$$E_{URLA}^{URLT}[C|\neg \text{star}] = \frac{(n-1)(n-2)(n-3)}{6(n-n^{4-n})} \tag{54}$$

and also

$$E_{URLA}^{URLT}[C|\neg \text{star}] = \frac{n-1}{3(n-n^{4-n})}|Q|_{linear}$$

$$\approx \frac{1}{3}|Q|_{linear}. \tag{55}$$

Adapting Eq. 28 to $E_{URLT}[h|\neg \text{star}]$, one obtains

$$E_{URLT}[h|\neg \text{star}] = \frac{n\left(E_{URLT}\left[\langle k^2 \rangle \middle| \neg \text{star}\right] - \langle k^2 \rangle_{linear}\right)}{(n-2)(n-3)}. \tag{56}$$

Note that

$$E_{URLT}\left[\langle k^2 \rangle \middle| \neg \text{star}\right] - \langle k^2 \rangle_{linear} = \frac{n-1}{1-n^{3-n}}\left(\frac{5n-6}{n^2} - n^{3-n}\right) - \left(4 - \frac{6}{n}\right)$$

$$= \frac{(n-2)(n-3)}{n}\frac{n^4 - n^n}{n^4 - n^{n+1}} \tag{57}$$

and then

$$E_{URLT}[h|\neg \text{star}] = \frac{n^{n-4} - 1}{n^{n-3} - 1}. \tag{58}$$

It is easy to see that

$$E_{URLT}[h|\neg \text{star}] \approx E_{URLT}[h] = \frac{1}{n} \tag{59}$$

for sufficiently large n. For numerical reasons, it is convenient to use Eq. 58 till $n = n^*$ and then replace the formula simply by $1/n$. n^* can be chosen as the largest value of n for which Eq. 58 does not produce numerical over-flows when calculating the powers. Such a critical value increases through the decomposition

$$E_{URLT}[h|\neg \text{star}] = a(n,1)a(n,-1), \tag{60}$$

with

$$a(n,x) = \frac{n^{\frac{n-4}{2}} + x}{n^{\frac{n-3}{2}} + x}. \tag{61}$$

All the corrected expectations that we have calculated in this section require $n \geq 4$ because there are no labelled trees with $n < 4$ such that they are not star trees.

Acknowledgements

References

[1] R. Cohen and S. Havlin. *Complex networks. Structure, robustness and function.* Cambridge University Press, Cambridge, UK, 2010.

[2] M. E. J. Newman. *Networks. An introduction.* Oxford University Press, Oxford, 2010.

[3] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, pages 167–256, 2003.

[4] E.A. Bender and E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A*, 24, 1978.

[5] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–180, 1995.

[6] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distribution and their applications. *Phys. Rev. E*, 64:026118, 2001.

[7] R. S. Milo, S. Shen-Orr, S.Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298:824827, 2002.

[8] R. Milo, N. Kashtan, S. Itzkovitz, M.E.J. Newman, and U. Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv preprint cond-mat/0312028*, 2003.

[9] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296:910913, 2002.

[10] A. C. C. Coolen, A. De Martino, and A. Annibale. Constrained Markovian dynamics of random graphs. *Journal of Statistical Physics*, 136(6):1035–1067, 2009.

[11] E.S. Roberts and A.C.C. Coolen. Unbiased degree-preserving randomization of directed binary networks. *Physical Review E 85*, 85:046103, 2012.

[12] Igor Mel'čuk. *Dependency Syntax: Theory and Practice*. State University of New York Press, 1988.

[13] Lucien Tesnière. *Éléments de syntaxe structurale*. Editions Klincksieck, 1959.

[14] David G. Hays. Dependency theory: A formalism and some observations. *Language*, 40:511–525, 1964.

[15] M. T. Gastner and M.E.J. Newman. The spatial structure of networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 49(2):247–252, 2006.

[16] M. Barthélemy and A. Flammini. Optimal traffic networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(07):L07002, 2006.

[17] R. Cohen and S. Havlin. *Distances in geographically embedded networks*, chapter Chapter 8, pages 88–90. Cambridge University Press, Cambridge, UK, 2010.

[18] Roger Levy, Evelina Fedorenko, Mara Breen, and Ted Gibson. The processing of extraposed structures in English. *Cognition*, 122(1):12 – 36, 2012.

[19] Yves Lecerf. Programme des conflits, modèle des conflits. *Bulletin bimestriel de l'ATALA*, 1:(4): 11–18, (5): 17–36, 1960.

[20] Ramon Ferrer-i-Cancho. Why do syntactic links not cross? *Europhysics Letters*, 76(6):1228–1235, 2006.

[21] Y. Albert Park and R. Levy. Minimal-length linearizations for mildly context-sensitive dependency trees. In *Proceedings of the 10th Annual Meeting of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT) conference*, pages 335–343, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[22] H. Liu. Dependency direction as a means of word-order typology: a method based on dependency treebanks. *Lingua*, 120(6):1567–1578, 2010.

[23] D. Gildea and D. Temperley. Do grammars minimize dependency length? *Cognitive Science*, 34(2):286–310, 2010.

[24] Carlos Gómez-Rodríguez and Joakim Nivre. Divisible transition systems and multiplanar dependency parsing. *Computational Linguistics*, 39(4):799–845, 2013.

[25] R. Ferrer-i-Cancho. Hubiness, length, crossings and their relationships in dependency trees. *Glottometrics*, 25:1–21, 2013.

[26] Ramon Ferrer-i-Cancho. A stronger null hypothesis for crossing dependencies. *Europhysics Letters*, 108:58003, 2014.

[27] R. Ferrer-i-Cancho. Random crossings in dependency trees. *Glottometrics*, 37:1–12, 2017.

[28] Ramon Ferrer-i-Cancho and Carlos Gómez-Rodríguez. Crossings as a side effect of dependency lengths. *Complexity*, 21:320–328, 2016.

[29] J. L. Esteban, R. Ferrer-i-Cancho, and C. Gómez-Rodríguez. The scaling of the minimum sum of edge lengths in uniformly random trees. *Journal of Statistical Mechanics*, page 063401, 2016.

[30] M. Noy. Enumeration of noncrossing trees on a circle. *Discrete Mathematics*, 180:301–313, 1998.

[31] R. Pastor-Satorras and A. Vespignani. *Evolution and Structure of Internet: A Statistical Physics Approach*. Cambridge University Press, 2004.

[32] Daniel Zeman, Ondřej Dušek, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, and Jan Hajič. HamleDT: Harmonized multi-language dependency treebank. *Language Resources and Evaluation*, 48(4):601–637, 2014.

[33] Rudolf Rosa, Jan Mašek, David Mareček, Martin Popel, Daniel Zeman, and Zdenek Žabokrtský. HamleDT 2.0: Thirty dependency treebanks stanfordized. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA).

[34] Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. Universal Stanford dependencies: a cross-linguistic typology. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA).

[35] Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, Magda Ševčíková-Razímová, and Zdeňka Urešová. Prague dependency treebank 2.0. CDROM CAT: LDC2006T01, ISBN 1-58563-370-4. Linguistic Data Consortium, 2006.

[36] R. Ferrer-i-Cancho. Euclidean distance between syntactically linked words. *Physical Review E*, 70:056135, 2004.

[37] R. Futrell, K. Mahowald, and E. Gibson. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341, 2015.

[38] R. Ferrer-i-Cancho. Non-crossing dependencies: least effort, not grammar. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, and B. Job, editors, *Towards a theoretical framework for analyzing complex linguistic networks*, pages 203–234. Springer, Berlin, 2016.

[39] H. Prüfer. Neuer Beweis eines Satzes über Permutationen. *Arch. Math. Phys*, 27:742–744, 1918.

[40] D. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Disc. Math.*, 3:450–465, 1990.

[41] A. Broder. Generating random spanning trees. In *Symp. Foundations of Computer Sci., IEEE*, pages 442–447, New York, 1989.

[42] R. Ferrer-i-Cancho and H. Liu. The risks of mixing dependency lengths from sequences of different length. *Glottotheory*, 5:143–155, 2014.

[43] D. Huff. *How to lie with statistics*. W. W. Norton and Company, New York, 1954.

[44] C. Gómez-Rodríguez and R. Ferrer-i-Cancho. The scarcity of crossing dependencies: a direct outcome of a specific constraint? *http://arxiv.org/abs/1601.03210*, 2016.

[45] Rudolf Rosa. Multi-source cross-lingual delexicalized parser transfer: Prague or stanford? In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 281–290, Uppsala, Sweden, 2015. Uppsala University.

[46] Ines Rehbein, Julius Steen, Bich-Ngoc Do, and Anette Frank. Universal Dependencies are hard to parse – or are they? In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pages 218–228, Pisa, Italy, 2017. University of Pisa.

[47] Carlos Gómez-Rodríguez. Restricted non-projectivity: Coverage vs. efficiency. *Computational Linguistics*, 42(4):809–817, December 2016.

[48] Nancy Ide and James Pustejovsky, editors. *Handbook of Linguistic Annotation*. Springer Netherlands, 1st edition, 2017.

[49] Roy Schwartz, Omri Abend, and Ari Rappoport. Learnability-based syntactic annotation design. In *Proceedings of COLING 2012*, pages 2405–2422. The COLING 2012 Organizing Committee, 2012.

[50] R. Ferrer-i-Cancho and F. Reina. Quantifying the semantic contribution of particles. *Journal of Quantitative Linguistics*, 9:35–47, 2002.

[51] A. Cayley. A theorem on trees. *Quart. J. Math*, 23:376–378, 1889.