

Divergence from Syntax to Linear Order in Ancient Greek Lexical Networks

Edoardo Maria Ponti

University of Pavia – Institute for Advanced Study (I.U.S.S.)
Strada Nuova, 65 – Piazza della Vittoria, 15
Pavia 27100, Italy

Abstract

All natural languages allegedly tend to minimize the length (and thus, the number of crossings) of syntactic dependencies by arranging words into an adequate linear order. Focusing on Ancient Greek as a case of study, this paper demonstrates that this tendency is far from constant. The method consists in inducing a pair of networks from a text representing a given diachronic variety, using lemmas as nodes and word relations (either syntactic dependency or linear adjacency) as arcs. Although the pair members share some topological properties (such as small-world effect, scale-freeness and disassortative mixing), they also diverge in some respects. By comparing the divergence in the Classic variety with that of the Late variety through Spectrum Analysis, a change is observed. This phenomenon may lead to rethinking how the cognitive complexity of language is measured and whether it is equivalent for different (varieties of) languages. In particular, this paper proposes the existence of a trade-off between dependency length minimization and pragmatic principles.

Introduction

The interaction between linear order and syntax is an ever-green topic in linguistics, encompassing constraints on word order of the main grammatical functions (Greenberg 1963) and rules for spelling out a given phrase structure (Kayne 1994). A new light may be cast on this interaction by network science, which arose as an attempt to model complex systems with a data-driven approach (Baronchelli et al. 2013). It has revealed thus far non-trivial facts in several domains, not least linguistics (Caldarelli 2007).

A network is a graph consisting of a set N of nodes (a synonym of vertices) and a set A of arcs (if directed) or edges (if undirected). A network can be induced from a corpus of texts by positing an equivalence between nodes and words, arcs and word relations. These relations can be syntactic in nature like the dependency of a child node from its parent in a tree representation. Moreover, they can be based also on the adjacency between a preceding word and a following word in a sentence (henceforth called co-occurrence). Neither of these is capable in itself to account for the whole range of linguistic phenomena: syntax dims the pragmatic

strategies of information structure. On the other hand, adjacent words are very often grammatically unrelated (Ferrer-i-Cancho, Solé, and Köhler 2004). Together, they give a more comprehensive picture.

The interest in objects like networks relies in the fact that they show global emergent properties, possibly different from the sum of those of local structures (i.e. single sentences). These properties represent general features of a language and are assumed to be pretty overlapping in networks based on dependency and co-occurrence (Solé et al. 2010). Nonetheless, this assumption has never been tested on pairs of networks induced from the same data source. The aim of this work is assessing whether the results of this test are consistent and, if a divergence is observed instead, whether it changes across different (varieties of) languages.

The rest of this work starts by providing a concise background on linguistic networks. Then it shows how the arrangement of words affects lengths and crossings of the syntactic dependencies of languages, and Ancient Greek in particular. The texts from which the networks are induced are then presented together with the results of their analysis. The degree of the divergence between dependency and co-occurrence networks across different varieties of Ancient Greek is measured through Spectrum Analysis. Finally, some conclusions are drawn from the results about the cognitive complexity of languages and the principles underpinning linear order arrangement.

Previous Works

The first attempt of inducing a network from a corpus exploited collocations (Ferrer-i-Cancho and Solé 2001), i.e. pairs of words appearing close to each other more often than random. Collocations approximated syntactic relations. Since then, networks based on this principle have been studied deeply and sophisticated filters have been developed to test the significance of collocations (Masucci and Rodgers 2009). Later, networks started being induced from treebanks, which are collections of texts already annotated with syntactic dependencies (Ferrer-i-Cancho, Solé, and Köhler 2004). These guarantee a better accuracy, without the need of pruning noise automatically. On the other hand, annotated data are more sparse than raw data.

These networks, independently from their induction process, were analysed according to some widely-used met-

rics. Both collocation and dependency networks were shown to display some shared properties. Firstly, they are small-world, i.e. compact and navigation-friendly. This effect is associated with both an average clustering coefficient higher than random by several orders of magnitude and an average shortest path length almost equivalent to random. The average clustering coefficient C measures the tendency of creating dense clusters, i.e. the probability that the neighbours of a node are themselves connected. Assume that c_n is the number of neighbours of a given node n , and e_n is the number of actually connected pairs among the neighbours of n . The ratio between e_n and the number of possible neighbour connections is averaged over every n :

$$C = \left\langle \frac{e_n}{c_n(c_n - 1)} \right\rangle \quad (1)$$

The value of C thus ranges from 0 (no possible pair instantiated) to 1 (all possible pairs instantiated). On the other hand, shortest path length D consists in the minimum number of arcs needed to reach one node from another averaged over every possible pair of arcs. If function $D_{min}(x, y)$ returns the minimum path length between nodes x and y , and the operator $\langle \dots \rangle$ expresses averaging, then:

$$D = \langle D_{min}(x, y) \rangle \quad (2)$$

Small-world networks are unaffected by random node deletion, but their properties change dramatically when a hub node (one of the most connected) is deleted (Albert, Jeong, and Barabási 2000). As for the comparison with random networks, these are generated through the following method (Erdős and Rényi 1961). Given a number of nodes $n = |N|$ and arcs $a = |A|$ from the original network, random pairs from n candidate nodes are linked until the required threshold of arcs a is reached.

Another renowned property is a power-law cumulative distribution of the degree $P(k)$: the degree k of a node equals to the number of its connections. If they are oriented, in- and out-degree are distinguished. $P(k)$ indicates the heterogeneity of a network (Barabási and Albert 1999) and abides by the following formula:

$$P(k) \propto k^{-\gamma} \quad (3)$$

Finally, linguistic networks display disassortative mixing, i.e. negative assortativity. This property captures the tendency of arcs to link nodes with similar degrees. It is measured as $\hat{k}_{nn}(k)$, the average degree of the neighbours of a node as a function of its degree¹.

Word Order

The network properties surveyed in the previous section mirror high-level properties of natural language. The current section points out that they dim at least a phenomenon that constitutes a parameter of variation across languages. Specifically, the amount of non-projective arcs, i.e. syntactic dependency arcs that intersect other arcs.

¹This has emerged as a valid alternative to Pearson’s correlation coefficient, which has statistical problems in large networks (Serrano, Boguñá, and Pastor-Satorras 2006).

Linear Arrangement Problem

The length of a syntactic relationship is defined as the Euclidean distance between its elements. The mean distance $\langle d \rangle$ over each connected pair in a sentence is proportional to the complexity of processing it sequentially. If $u \sim v$ is a pair of nodes connected by an arc, $\pi(u)$ is a function returning the position in linear order of node u , and n is the number of tokens in the sentence, then:

$$\langle d \rangle = \frac{\sum_{u \sim v} |\pi(v) - \pi(u)|}{(n - 1)} \quad (4)$$

The word orderings of the natural languages stem from the choice of an optimal solution to the problem of minimizing $\langle d \rangle$, i.e. the so-called linear arrangement problem (Ferrer-i-Cancho 2006). Among all possible linear orderings, only the subset satisfying this requirement is pervasively attested. In turn, the low frequency of the crossings of syntactic dependency arcs is a side effect of this optimization, rather than originating from an independent principle. This is because a crossing becomes more likely with longer arcs (Gómez-Rodríguez and Ferrer-i-Cancho 2016).

The ban against the presence of crossings is called projectivity, which is defined as follows: any dependency subtree must cover a contiguous region of a sentence (Marcus 1965). Assume that $i \rightarrow j$ stands for a dependency between nodes i and j and instead i, j stands for the array of elements delimited by these nodes (included). $Subtree_i$ is the set of nodes that can be reached following oriented arcs starting from node i with an arbitrary number of steps. Then the projectivity requirement can be expressed as in this formula (Havelka 2007):

$$i \rightarrow j \wedge v \in (i, j) \Rightarrow v \in Subtree_i \quad (5)$$

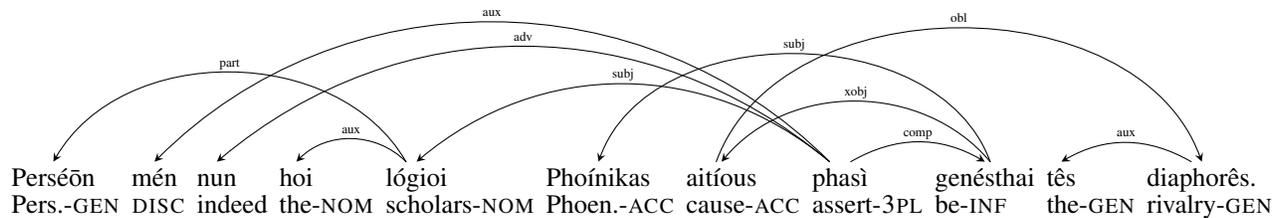
Violations of this constraint are nonetheless ubiquitous. If it does not hold, then v is said to be “in a gap”. An arc covering at least a node in a gap is non-projective. The percentage of these arcs is a common measure of how much non-projectivity is widespread in a certain language.

Non-projectivity in Ancient Greek

Ancient Indo-European languages have more non-projective arcs compared to their modern descendants. Among the former group, the most striking rate shows up in Ancient Greek. In the Ancient Greek Treebank (Bamman, Mambrini, and Crane 2009), which contains texts from Homer to Plato, the percentage of non-projective arcs is 15.15%. In a sample group of modern Indo-European languages, a range from 1.37% (Portuguese) to 5.90% (Dutch) is reported instead (Mambrini and Passarotti 2013).

The authors lead back this idiosyncrasy to some linguistic phenomena. In particular, clitic postpositive particles tend to be in a gap. As they are prosodically non-autonomous, they occur in some fixed positions, most notably the second one (Wackernagel’s law). The second source of non-projectivity is the drift of modifiers apart from their head.

Consider Figure 1 as an example: the discourse particles *mén* and *nun* crowd around the second position and are in a gap, as they do not belong to the subtree including *lógiou*



‘Indeed, the Persian scholars assert the Phoenicians to be responsible for the rivalry’ Herodotus, *Histories* I 1.1.

Figure 1: A syntactic dependency tree, with labeled arcs above and the morphological glosses below. Abbreviations: *part* = partitive, *aux* = auxiliary word, *adv* = adverbial, *subj* = subject, *xobj* = predicative object, *comp* = complement, *obl* = oblique.

‘scholars’ and *Perséōn* ‘of the Persians’. Moreover, a dependent, the subject *Phoínikas* ‘the Phoenicians’, is distanced from its head, the subordinate verb *genésthai* ‘to be’, by the predicative object *aítious* ‘responsible’ and the main verb *phasì* ‘they assert’.

Nonetheless, the rate of non-projective arcs tapered off in later varieties of Ancient Greek (as well as in all Ancient Indo-European languages). This is possibly due to both a different clitic positioning that weakened the Wackernagel’s law (Taylor 1990) and a grammaticalization of continuous constituents, which merged in the same phrase a head with its modifier (Devine and Stephens 2000). In fact, a trend towards a more fixed word order increased the rate of adjacent head-modifier pairs (Gulordava and Merlo 2015). For these reasons, the Classic and Late varieties of Ancient Greek, which represent the beginning and the middle stages of this evolution, fit well as a case of study. They are as similar as possible, but they differ in a parameter affecting deeply word order, namely the degree of non-projectivity.

Data and Method

The data are obtained from the collection of dependency treebanks developed through the PROIEL (Pragmatic Resources in Old Indo-European Languages) project (Haug and Jøhndal 2008). Two of them were selected: one contains the *Histories* by Herodotus (440-429 B.C.) and the other the *Greek New Testament* (49-150 ca. A.D.): they represent Classic and Late varieties, respectively². Both were levelled off to the same amount of tokens, namely 67247.

The percentage of non-projective arcs amounts to 9.27% in the Classic variety, against 2.65% in the Late variety³. The basic intuition is that this makes word order diverge from syntax more in the former than in the latter. Of course, there is an inherent difference between dependencies and co-occurrences. In a sentence, every word that does not lie in the tail has exactly an adjacent following word. On the other hand, the number of its syntactically dependent words can

range from zero to the length of the sentence minus one. The role of non-projectivity is scattering even more dependent words far from the position adjacent to their head. In fact, the percentage of arcs that overlap in adjacency and dependency (i.e. the number of relations with Euclidean distance 1) is 38.04% in the Classic variety, and 46.51% in the Late variety.

From each of these two treebanks, a dependency networks was induced through a previously devised method (Ferrerri-Cancho et al. 2007): a pair of words is connected if there is at least a syntactic relation between them in a sentence. However, the method used in this paper differs in some respects: it is based on lemmas rather than forms to reduce data sparseness. Furthermore, the graph is oriented to distinguish head from dependent nodes. To my knowledge, a global dependency network of an ancient Indo-European language (Latin) has been investigated only once (Passarotti 2014).

Secondly, the induction process was repeated, this time based on co-occurrence. Two nodes in the networks are linked if they are adjacent at least once in a sentence. This method, contrary to those based on collocations, is not filtered: its aim is not approximating syntax, but rather mirroring exactly linear order. In order to be comparable, co-occurrence networks are lemma-based and oriented, too. The latter condition allows to distinguish a following from a preceding node. Basic information about the four induced networks is summed up in Table 1⁴.

Variety	Relation	Name	Nodes	Edges
classic	dependency	CD	5378	34244
classic	adjacency	CA	5398	34076
late	dependency	LD	3005	20695
late	adjacency	LA	3025	20788

Table 1: Details about the Ancient Greek networks.

Results

The first part of this section shows that co-occurrence and dependency networks mainly share their topological properties, although some differences obtain. In particular, the

⁴Note that the values refer to the largest connected component.

²Although they belong to different genres and are limited to a single book, they are both written in prose. Finding representative and comparable resources for ancient languages is problematic.

³These values are affected also by the average sentence length, namely 20.29% against 15.84%, respectively. In effect, longer sentences favour higher non-projectivity.

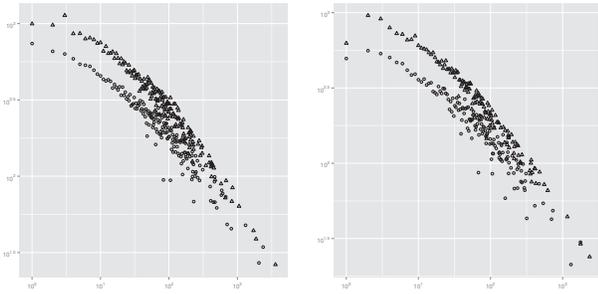


Figure 2: Average neighbour degree as a function of the degree for Classic (left) and Late (right) varieties. The x-axis represents degree, whereas y-axis average neighbour degree. Adjacency is plotted as triangles, dependency as circles.

former surpasses the latter in small-world effect and disassortative mixing. The second part instead will be focused on the relationship between a pair of networks for a given variety, relying upon spectrum analysis. This allows to provide a measure for the possible divergence and compare divergences themselves.

Topological Properties

The metrics related to the small-world effect were computed through the R package *igraph* (Csardi and Nepusz 2006). Table 2 shows the values of the average clustering coefficient and shortest path length of the networks in Table 1 and of their corresponding Erdős-Rényi random graphs. It emerges that all the four networks are small-world, since $C \gg C_{random}$ and $D \approx D_{random}$. Furthermore, it is noteworthy that in co-occurrence networks the small-world effect, and hence compactness, are stronger: intuitively, adjacency allows more alternative follow-ups for a given node than dependency.

Name	C	D	C_{random}	D_{random}
CD	0.333	3.368	0.002	4.875
CA	0.457	3.064	0.002	4.875
LD	0.432	3.410	0.005	4.372
LA	0.560	2.865	0.005	4.372

Table 2: Clustering coefficients and shortest path lengths.

As for cumulative degree distribution, the nodes with a high degree, called hubs, are generally verbs in dependency networks. This is because they dominate directly all their arguments and modifiers, the so-called full valency, in their sentence trees (Čech, Mačutek, and Žabokrtský 2011). On the other hand, hubs are function words in co-occurrence networks due to their frequency (Ferrer-i-Cancho and Solé 2001). The negative exponent of the power law in Formula 3 fitting the cumulative degree distribution plot is 1.946 for CD and 1.906 for CA, 1.981 for LD and 1.894 for LA: all of them are pretty overlapping. The networks appear to be scale-invariant, but they miss the expected range of the exponent $2 < \gamma < 3$ (Solé et al. 2010). Perhaps, what flattens

the slope of the power law is the downscaling of function words and their bridging role in languages with a rich morphology: they are undermined by inflectional suffixes, but these are deleted in lemma-based networks.

Finally, consider the scatter plot in Figure 2. Although it cannot be fit by a power law, average degree of the neighbours of a node clearly drops as the node degree arises. Therefore, both co-occurrence and dependency networks are disassortative. Their data points are completely separated, though, especially with a low degree. As a consequence, co-occurrence networks are more disassortative.

Spectrum Analysis

A network consisting in a set of nodes N and arcs A can be conceived as a binary adjacency matrix \mathbf{A} . This is a matrix with $|N|$ rows and $|N|$ columns, where the cell (i, j) is filled with 1 if there is an arc linking i and j in A , and filled with 0 otherwise. Spectrum analysis is based on the eigenvalues of this matrix. λ is an eigenvalue for \mathbf{A} if there exists an $|N|$ -dimensional vector x that satisfies this equation:

$$\mathbf{A}x = \lambda x \quad (6)$$

The non-zero vector x is said to be an eigenvector associated to its eigenvalue λ . The spectrum is a density distribution of the set $\Lambda \{\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{|N|-1}\}$ (whose cardinality is $|N|$) and the multiplicities of its members (roughly, the times they repeat).

Spectrum analysis unravels global patterns that remain hidden to the metrics mentioned thus far (Choudhury, Chatterjee, and Mukherjee 2010). Co-occurrence networks induced from real texts are distinguishable from small-world, scale-free, and disassortative networks generated randomly through the Dorogovtsev-Mendes growth model (Dorogovtsev, Mendes, and Samukhin 2000) only because of their spectra. In particular, in the former the eigenvalues are more densely concentrated around 0.

According to the authors, this difference stems from the fact that the probability a word co-occurs with any other is not the pure outcome of its frequency. Rather, words have precise constraints on their neighbourhood. In particular, elements belonging to the same part of speech tend to have pretty overlapping sets of neighbours. This pattern results into a high number of similar rows in the matrix. As the rank is the measure of how many linearly independent rows (or, equivalently, columns) are found in a matrix, density under 0 means low rank and hence it entails the regularity of some combinations.

The four networks induced in this paper were treated for the current purpose as unoriented, so that each \mathbf{A} be symmetric, and its eigenvalues fall in the domain of real numbers. Eigenvalues and multiplicities were calculated for each and the resulting density distribution is plotted in Figure 3. Whereas the dependency-based spectra are quite similar in both their peak and slope, the co-occurrence-based spectra vary to a greater extent. The co-occurrence-based spectrum diverges from the respective dependency-based one more in the Late variety than in the Classic variety, where the values are less shrank around 0. The co-occurrence-based matrix

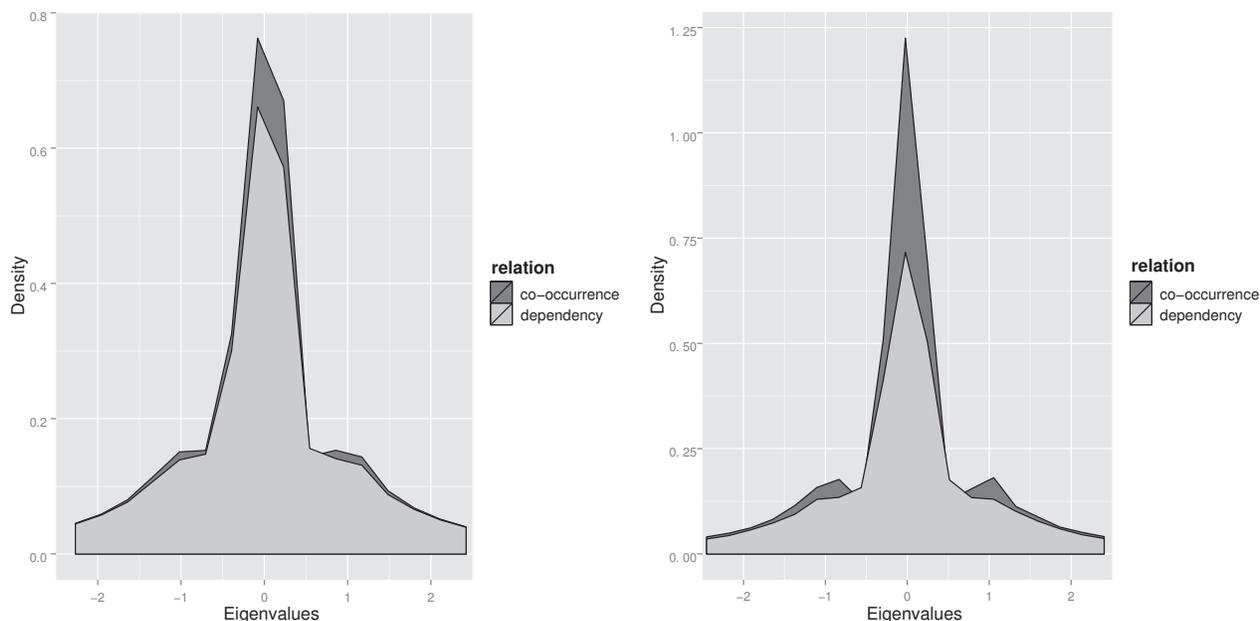


Figure 3: Density distributions of the spectra of Classic (left) and Late (right) varieties of Ancient Greek, zoomed around 0.

of the Classic variety in fact is expected to be less regular, because of the higher freedom of its word order.

Discussion

According to the results, the properties of dependency networks are stable, confirming the hypothesis of their universality (Ferrer-i-Cancho et al. 2007). Co-occurrence network properties are prone to variation instead. Assuming that language networks have a cognitive reality, and that both syntactic dependencies and linear order are relevant for language comprehension and generation, then it is tempting to raise a question: are some languages more complex than others (Sampson, Gil, and Trudgill 2009)?

The results reveal two facets, both at the global (network) and local (sentences) level. Starting from the former, consider the Classic variety of Ancient Greek and its co-occurrence network. On the one hand, the low amount of eigenvalues around 0 points towards a lack of regularity. On the other hand, the strong small-world effect makes the network fit for computability: a high clustering coefficient enables a fast local processing, and a low average shortest path length guarantees an efficient global integration (Deco et al. 2008). Both these effects originate from the the higher average dependency length of the Classic variety compared to the Late variety, which results into a freer word order and a high number of crossings (Gómez-Rodríguez and Ferrer-i-Cancho 2016).

In fact, as for the local level, dependency length minimization plays a minor role in Classic Ancient Greek (although the average dependency length is still lower than random (Gulordava and Merlo 2015)). The reason lying behind this apparent anomaly is that word arrangement is constrained by multiple factors. For example, it has been

shown that languages do not favour dependency length minimization during the early stages of their evolution, selecting subject-object-verb order. Later, this is transformed into subject-verb-object order, which is optimal for minimization (Ferrer-i-Cancho 2014). In particular, the author maintains that the factor setting off against dependency length minimization (conceived as on-line memory minimization) is the maximum predictability of the verb (by its final position).

Nonetheless, predictability cannot account for Ancient Greek linear order because its positions are not fixed. Instead, the explanation may be led back to iconicity (Haiman 1980). Iconicity ensures an isomorphism between syntactic and semantic structures. Thus, semantically related elements (such as heads and dependents) are linearly close: this improves regularity and results into dependency length minimization. On the other hand, initial positions correlate with old information, and vice versa final positions with new information. Moreover, what is uppermost (most relevant) in the speaker's mind is expressed with the highest priority (Croft 2002). The last two principles are candidate to be responsible for the Classic Ancient Greek behaviour. Future studies may verify this hypothesis and possibly integrate them into the framework of the linear arrangement problem.

Conclusion

Considering networks induced from the same Ancient Greek texts, this paper demonstrated that dependency networks display constant behaviours, whereas those of co-occurrence networks vary. This leads to different extents of divergence between the two across different varieties of language. The significance of these results is that a comparison of this sort has been accomplished for the first time: the related variation was thus far unobserved. This work proposed that the vari-

ation stems from competing motivations in linear arrangement, i.e. dependency length minimization against information and relevance principles. All of them may emerge from a common reason, i.e. that sentence position is iconic with respect to semantics and pragmatics.

Future enhancements possibly include polishing the network representations. Lemmas should be substituted with forms as network nodes. Inflectional morphology is pervasive in richly fusive languages such as Ancient Greek, and this would enable a fair comparison with languages endowed with independent function words. Furthermore, weighting the arcs would make complex networks more reliable (Barrat et al. 2004).

Acknowledgements

Thanks to the Institute for Advanced Study of Pavia for supporting my participation to the conference and to Marco Passarotti for his precious advices.

References

- Albert, R.; Jeong, H.; and Barabási, A.-L. 2000. Error and attack tolerance of complex networks. *Nature* 406(6794):378–382.
- Bamman, D.; Mambrini, F.; and Crane, G. 2009. An ownership model of annotation: The Ancient Greek dependency treebank. In *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT 8)*, 5–15.
- Barabási, A.-L., and Albert, R. 1999. Emergence of scaling in random networks. *Science* 286(5439):509–512.
- Baronchelli, A.; Ferrer-i-Cancho, R.; Pastor Satorras, R.; Chater, N.; and Christiansen, M. H. 2013. Networks in cognitive science. *Trends in cognitive sciences* 17(7):348–360.
- Barrat, A.; Barthelemy, M.; Pastor-Satorras, R.; and Vespignani, A. 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* 101(11):3747–3752.
- Caldarelli, G. 2007. Scale-free networks: complex webs in nature and technology. *OUP Catalogue*.
- Čech, R.; Mačutek, J.; and Žabokrtský, Z. 2011. The role of syntax in complex networks: Local and global importance of verbs in a syntactic dependency network. *Physica A: Statistical Mechanics and its Applications* 390(20):3614–3623.
- Choudhury, M.; Chatterjee, D.; and Mukherjee, A. 2010. Global topology of word co-occurrence networks: Beyond the two-regime power-law. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 162–170.
- Croft, W. 2002. *Typology and universals*. Cambridge University Press.
- Csardi, G., and Nepusz, T. 2006. The igraph software package for complex network research. *InterJournal, Complex Systems* 1695(5):1–9.
- Deco, G.; Jirsa, V. K.; Robinson, P. A.; Breakspear, M.; and Friston, K. 2008. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Computational Biology* 4(8).
- Devine, A. M., and Stephens, L. D. 2000. *Discontinuous syntax: hyperbaton in Greek*. Oxford University Press.
- Dorogovtsev, S. N.; Mendes, J. F. F.; and Samukhin, A. N. 2000. Structure of growing networks with preferential linking. *Physical review letters* 85(21):4633.
- Erdős, P., and Rényi, A. 1961. On the strength of connectedness of a random graph. *Acta Mathematica Hungarica* 12(1-2):261–267.
- Ferrer-i-Cancho, R., and Solé, R. V. 2001. The small world of human language. *Proceedings of the Royal Society of London B: Biological Sciences* 268(1482):2261–2265.
- Ferrer-i-Cancho, R.; Mehler, A.; Pustynnikov, O.; and Díaz-Guilera, A. 2007. Correlations in the organization of large-scale syntactic dependency networks. *TextGraphs-2: Graph-Based Algorithms for Natural Language Processing* 65–72.
- Ferrer-i-Cancho, R.; Solé, R. V.; and Köhler, R. 2004. Patterns in syntactic dependency networks. *Physical Review E* 69(5):051915.
- Ferrer-i-Cancho, R. 2006. Why do syntactic links not cross? *Europhysics Letters* 76(6):1228–1235.
- Ferrer-i-Cancho, R. 2014. Why might sov be initially preferred and then lost or recovered? a theoretical framework. In *The Evolution of Language-Proceedings of the 10th International Conference (EVOLANG’10)*, 66–73.
- Gómez-Rodríguez, C., and Ferrer-i-Cancho, R. 2016. The scarcity of crossing dependencies: a direct outcome of a specific constraint? *arXiv preprint arXiv:1601.03210*.
- Greenberg, J. H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, J. H., ed., *Universals of language*, volume 2. MIT Press. 73–113.
- Gulordava, K., and Merlo, P. 2015. Diachronic trends in word order freedom and dependency length in dependency-annotated corpora of latin and ancient greek. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 121–130.
- Haiman, J. 1980. The iconicity of grammar: isomorphism and motivation. *Language* 515–540.
- Haug, D. T., and Jøhndal, M. 2008. Creating a parallel treebank of the old Indo-European Bible translations. In *Proceedings of the Language Technology for Cultural Heritage Data Workshop (LaTeCH 2008), Marrakech, Morocco, 1st June 2008*, 27–34.
- Havelka, J. 2007. Beyond projectivity: Multilingual evaluation of constraints and measures on non-projective structures. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, volume 45, 608–615.
- Kayne, R. S. 1994. *The antisymmetry of syntax*. MIT Press.
- Mambrini, F., and Passarotti, M. 2013. Non-projectivity in the Ancient Greek Dependency Treebank. *DepLing 2013* 177.
- Marcus, S. 1965. Sur la notion de projectivité. *Mathematical Logic Quarterly* 11(2):181–192.
- Masucci, A., and Rodgers, G. 2009. Differences between normal and shuffled texts: structural properties of weighted networks. *Advances in Complex Systems* 12(01):113–129.
- Passarotti, M. 2014. The importance of being *sum*. Network analysis of a latin dependency treebank. In *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014*. Pisa University Press.
- Sampson, G.; Gil, D.; and Trudgill, P. 2009. *Language complexity as an evolving variable*, volume 13. Oxford University Press.
- Serrano, M. Á.; Boguñá, M.; and Pastor-Satorras, R. 2006. Correlations in weighted networks. *Physical Review E* 74(5):055101.
- Solé, R. V.; Corominas-Murtra, B.; Valverde, S.; and Steels, L. 2010. Language networks: their structure, function, and evolution. *Complexity* 15(6):20–26.
- Taylor, A. 1990. *Clitics and configurationality in Ancient Greek*. Ph.D. Dissertation, University of Pennsylvania.