

A Treebank-based Approach to the *Suprema Constructio* in Dante’s Latin Works

Flavio Massimiliano Cecchini, Giulia Pedonese

Università Cattolica del Sacro Cuore di Milano
Largo Gemelli 1, 20123 Milano (MI), Italy
{flavio.cecchini,giulia.pedonese}@unicatt.it

Abstract

This paper aims to apply a corpus-driven approach to Dante Alighieri’s Latin works using UDante, a treebank based on *Dante Search* and part of the Universal Dependencies project. We present a method based on the notion of barycentre applied to a dependency tree as a way to calculate the “syntactic balance” of a sentence. Its application to Dante’s Latin works shows its potential in analysing the style of an author, and contributes to the interpretation of the *suprema constructio* mentioned in *DVE* II vi 7 as a well balanced syntactic pattern modeled on Latin literary writing.

Keywords: Dante, treebanks, stylometrics

1. Introduction and Motivation

Dante Alighieri (1265–1321) is known as the author of the *Divine Comedy*, a poem describing the author’s journey in the afterlife. This is legitimately considered his masterpiece, and its immediate success contributed to the creation of the Italian language (De Mauro and Allasia, 2005).

However, Dante was also a bilingual author writing verse and prose in Italian (*Rhymes*, *Vita Nova*, *The Banquet*, *The Divine Comedy*) and in Latin (*De vulgari eloquentia*, *Monarchia*, *Epistles*, *Eclogues*, *Questio de aqua et terra*).¹ This was not uncommon in the 14th Century, when Latin was the standard language for official writing and Italian was used for specific literary traditions such as the Sicilian love poetry, practical documents and informal communication. In his works, Dante makes an effort to promote the Italian language to a higher level by modeling it on Latin. For this reason, studying Dante’s Latin could shed light on this creative process.

However, the tradition of Dante Studies lacks a systematic analysis of this side of Dante’s production. The gap has been pointed out as a major issue (Curtius, 1948; Paratore, 1965; Brugnoli, 1965; Basile and Brugnoli, 1971) and is complicated by problems of attribution like those persisting around the *Epistle* XIII and the *Questio*.²

So far, the majority of linguistic studies and language resources have been focusing on Dante’s Italian works, but a pivotal role has been played by *Dante Search*, a corpus offering the complete grammatical annotation of Dante’s works and a syntactic annotation limited to

his Italian production (Tavoni, 2011). Sporadic studies on the lexicon of Dante’s Latin works are available, but only to highlight Dante’s linguistic peculiarities through new formations and *hapax legomena*, especially those of the *DVE*, collected in a glossary attached to Aristide Marigo’s critical edition (Alighieri, 1938).

More recent efforts have been made by the new *Vocabolario Dantesco Latino* (Albanese et al., 2019), a dictionary whose goal is to provide the first systematic study of Dante’s Latin lexicon through the extensive use of Classical and Medieval Latin corpora. However, the project has only just been started with the publication of 119 entries and its major concern, as a dictionary, is not Dante’s Latin syntax, although this aspect is often taken into account when relevant from the lexicographic standpoint. So far, the only complete lexical resource available for Dante’s Latin is UDante, a treebank based on *Dante Search* and developed according to Universal Dependencies’ guidelines (Cecchini et al., 2020b) and recently linked to the LiLa Knowledge Base (Passarotti et al., 2021) (see §2).

However, no one has ever attempted a description of Dante’s Latin syntax despite the importance of such analysis in the relationship between the author and his sources. This is particularly relevant since Dante shows a unique theory of syntax in *DVE* II vi 7: here he calls *suprema constructio* the perfect syntactic structure taken from both Latin poetry and prose in order to write poetry in Italian in the highest possible style. Given the lack of a formal theory of syntax in ancient and Medieval times, Dante did not have the tools to articulate his intuition any further, and the definition of this syntactic pattern is still unclear. However, the *suprema constructio*, as Dante describes it, is applicable to both Latin and Italian regardless of the distinction between poetry and prose, thus allowing an interlinguistic approach.

The first corpus-driven study applied to Dante’s syntax (Tavoni and Chersoni, 2013) is an attempt to formally

¹For which the abbreviations used in the following are respectively *Mon*, *DVE*, *Egl*, *Epi*, *Que*.

²To cite only the most recent studies, a new attempt to verify the attribution of the epistle through machine learning has been made by Corbara et al. (2020). As for the *Questio*, Fioravanti (2017) points out that one of the used arguments was not known until after 1320, the date of its discussion.

describe the *suprema constructio* by extending a simplified version of the syntactic annotation of Dante’s Italian works included in *Dante Search*. The interpretation resulting from Tavoni and Chersoni’s study is that this pattern is characterized by a relatively low depth and symmetry, linking Dante’s definition of *suprema constructio* to the idea of balance. Although their study is mainly focused on romance languages, the authors also point out that the analysis should be extended to the Latin authors mentioned in DVE II *vi* 7 as examples of style (Vergil, Ovid’s *Metamorphoses*, Statius, Lucan, Livy, Pliny,³ Frontinus and Paulus Orosius), but to this day the available annotated language resources only allow to create a small portion of such a corpus, at least in the context of Universal Dependencies (see §2). This paper’s aim is to develop a topological method in order to ground the stylistic interpretation of texts into the linguistic and mathematical representation of the dependency tree, as detailed in §3. The paper also presents (§4) a first application of this corpus-driven method to the analysis of Dante’s Latin syntax to assess if the notion of *suprema constructio* could have interfered with his Latin writing, before concluding (§5). All data and some scripts used in this paper are made available at the GitHub repository of one of the authors.⁴

2. Data

Our investigation is primarily conducted on data from the latest (v2.10) version of UDante⁵ (Tavoni, 2011; Cecchini et al., 2020b), itself a treebank part of the Universal Dependency (UD) project⁶ (de Marneffe et al., 2021) and including all five Latin works by Dante, as mentioned in §1.

Unfortunately, a survey of the availability of the Classical works cited by Dante as examples of “good (Latin) style” (see §1) in the same UD framework leads to unsatisfactory results. Only the UD Perseus treebank⁷ (Bamman and Crane, 2011) and UD PROIEL⁸ (Eckhoff et al., 2018) contain Classical texts, and among these only the *Metamorphoses* by Ovid and the *Aeneid* by Vergil in Perseus are of interest to the current work, and then only respectively book I and book VI are present. Further, both UD Perseus and UD PROIEL have “neglected” status as of UD v2.10,⁹ meaning that their

annotation quality is problematic in UD terms;¹⁰ further still, the aforementioned texts are incomplete, as many sentences are missing,¹¹ so that we are finally left with only 68 sentences for the *Aeneid* and 183 for the *Metamorphoses*, some of which are themselves only snippets of more complex periods.¹² In the end, their scarcity and problematic annotation quality mean that we have to refrain from using these data in our investigation.

An attempt to obtain a more ample data set by using the UDPipe POS-tagger (Straka et al., 2016) on complete raw texts has also not yielded any acceptable results. To test the potential of this approach, we sketch an evaluation of a UDPipe model trained on all UD Latin treebanks on the first ten sentences it identifies on the raw texts of respectively book I of the *Metamorphoses* and book VI of the *Aeneid*, as taken from the Perseus Digital Library.¹³ The gold standard is created as the manual correction of the UDPipe output by the two authors, following the latest standards for the annotation of Latin in UD, as exemplified by the UDante treebank. This results in two test sets of 234 and 199 tokens each. Since we are mostly interested in the structure of dependency trees rather than in the specific labels of dependency relations (see §3.1), we compute the *unlabeled attachment score* (UAS),¹⁴ which ends up being an extremely low 40.2% in both cases (labeled attachment score is at 33.3% and 31.7% respectively). Despite the very small test sets, these scores, corroborated by further manual inspection, are evidence for a still unreliable automated parsing on which we cannot reasonably base our study: therefore, we have to stick only to the existing active UD treebanks for Latin, which are, besides UDante, the IT-TB (Passarotti, 2019) and LLTC (Cecchini et al., 2020a). In fact, despite such a negative appraisal for our specific case (and observing that Clas-

mff.cuni.cz/udvalidator/cgi-bin/unidep/validation-report.pl.

¹⁰Especially for what concerns parts of speech, morphological features and dependency relation, while the overall tree structures can be considered mostly sound. In fact, the UD version of these treebanks derives from a structurally reliable automated conversion between the original, manually annotated format as described by Bamman et al. (2007); also refer to (Cecchini et al., 2020a, §2) for details about this conversion process.

¹¹Compare for example book I of the *Metamorphoses* in the treebank (sentences beginning with phi0959) and the original at <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.02.0029:book=1>.

¹²In Perseus, sentences are split also at “weak” punctuation marks such as : and ;, differently than in UDante. This means among other things that co-ordinating and paratactical constructions will be underrepresented; see §3.3 and §4.

¹³<https://www.perseus.tufts.edu>

¹⁴The employed software is MaltEval (Nilsson and Nivre, 2008). Some formal adjustments of the CoNLL-U files are needed to take into account different tokenisations between automated output and manual gold standard.

³It is still unclear if Dante refers to Pliny the Elder or Pliny the Younger, see Tavoni’s commentary in (Alighieri, 2011, pp. 1454–1455).

⁴<https://github.com/Stormur/DanteSupremaConstructio>

⁵https://github.com/UniversalDependencies/UD_Latin-UDante

⁶<https://universaldependencies.org/>

⁷https://github.com/UniversalDependencies/UD_Latin-Perseus

⁸https://github.com/UniversalDependencies/UD_Latin-PROIEL

⁹See UD validation page at <http://quest.ms>.

sical Latin is underrepresented in UD), the structures described in §3 and the discussion framework in §4 are general enough that they can be pursued for any other UD treebank, and, even more generally, can be adapted to any set of treebanks, provided that they share an annotation formalism, so as to obtain meaningful comparisons. A big part of the difficulties encountered by UDPipe is most probably due to the great differences in lexicon and style between poetry on the one side, and a prevalence of treatise prose in UD Latin treebanks on the other; see e. g. the discussion in (Ponti and Passarotti, 2016, §7).

3. Linguistico-mathematical Background

The syntactic representation of a sentence following a dependency paradigm such as in UD (as opposed to a constituency or phrase approach; see (Osborne, 2019, §2)) is usually called a **dependency tree**, and, at a mathematical level, is defined (Havelka, 2007, §1) as a graph possessing an ordering of the nodes (corresponding to the linear order of the words) and being a) **directed**, i. e. each edge has a fixed orientation from one end to the other, and b) **rooted**, i. e. each node has at most one parent and there is one and only one node, the **root**, from which all other nodes can be reached. The corresponding, and motivating, linguistic interpretation of the root node (represented by the homonymous relation `root` in UD) is that of the most dominant element in the clause: usually, in non-elliptic clauses, the predicate (most frequently expressed by a verb, i. e. an element with part of speech `VERB` in UD), which determines the syntactic (argumental) structure and lexical composition of the utterance. However, on more mathematical bases, also other kinds of *centrality* notions can be defined and exploited (§3.1). Moreover, the interplay between the two dimensions of linear word order on one part and underlying non-linear syntactic structure on the other defines the notion of *non projectivity* (§3.2), which supplies a further interpretation key to the stylistic analysis of a text (see §4).

3.1. Roots and Barycentres

While the specific mathematical structure of dependency trees is chosen to represent linguistic structure also, but not exclusively, in view of given theoretical-linguistic assumptions, we can try to reverse this perspective and apply purely mathematical instruments on it to help linguistic investigations. In this sense, to pursue the stylistic analysis of Dante’s work, we introduce the notion of **barycentre** (or median) of an undirected graph. This is defined in general (West, 2001, §2.1.55) as the set of nodes $\mathcal{B} \subset N$ in a graph $G = (N, E)$ such that for any node $b \in \mathcal{B}$ its so-called **total distance** $t(b) = \sum_{n \in N} d(b, n)$, i. e. the sum of the distances¹⁵ from b to each other node, is minimal in the

¹⁵The distance is defined as the number of edges on the shortest path(s) between two nodes; in a tree, given it is acyclical, the shortest path between two nodes is unique

graph G . To compute the barycentre on a dependency tree, we must first consider its underlying undirected graph; then, the properties of trees assure us that the barycentre will always consist of either a single node or an edge, i. e. two adjacent nodes (Koschützki et al., 2005, §3.3.4). Now, in a dependency tree the root does not necessarily lie in the barycentre: we illustrate this in Figure 1¹⁶ with a short sentence from the UDante corpus. Here, the predicate, the finite verb form *videtur*, is the root and governs a clausal subject (`csubj`), headed by the verb form *exaltatum*, which is the barycentre: a quick computation yields indeed a total distance of 7 for *exaltatum*, while of 10 for *videtur*, of 15 for *autem* and of 11 for all other nodes (see §3.3 for the details).

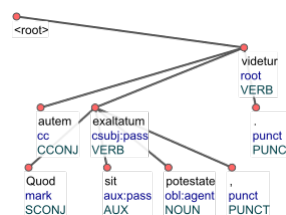


Figure 1: Sentence DVE-186 with $\Delta = 1$.

The distance Δ between the root and the barycentre has only a lower bound depending on the **depth** of the dependency tree, i. e. the maximum distance from the root to any other node: in Figure 1 the depth is 2, so Δ can be at most 1 here. In general, if the depth is k , Δ can vary between 0 (i. e. root and barycentre coincide) and $k - 1$ (the barycentre cannot be a leaf node): for each of these values, it is possible to construct a dependency tree such that it attains that value,¹⁷ so there are no other particular restrictions on Δ , apart those intrinsic in natural languages (cf. §4).

The linguistic interpretation that we associate to the barycentre and Δ is that of **syntactic balance**: the barycentre can be seen as the “main branching point” of the dependency tree, where the sentence is developed and expanded the most. A Δ of 0 implies that the arguments of the sentence are distributed (in a syntactic sense) “harmoniously”, or symmetrically, with respect to the root, while greater values mean that the sentence “hinges” more heavily on a particular subordinated element and that it expands this more than others. This can be observed in Figure 1, where *videtur* ‘it appears that’ is seen to function just as a frame for the actual assertion *exaltatum sit potestate* ‘it is exalted by power’ (through the link *quod* ‘that’). We propose this Δ as a more universally suited measure of syntactic balance

(West, 2001, §2.1.4). The distance $d(b, b)$ of a node from itself is zero.

¹⁶Produced by means of the CoNLL-U Viewer online tool at https://universaldependencies.org/conllu_viewer.html.

¹⁷The proof is rather trivial, but we unfortunately have no space to show it in this paper.

than the ASM (“asymmetry index”) presented by Tavoni and Chersoni (2013, §5), by which it is ultimately inspired.

The problem of ASM is that of being based on a too literal notion of geometric, symmetric centrality with respect to the root: this might fit to a verb-medial order as found in Medieval Romance varieties, but is no longer applicable to a more variable word order with verb-final tendencies (at least in literary language) like that of Latin (cf. the diachronic perspective in (Ledgeway, 2012, §3.3)). In a typological perspective, we thus need to turn to a topological (i. e. based on the relative, not absolute, positions of the nodes) notion, like the barycentre, to take into account the syntactic structure of a sentence with no assumptions on its actual word order (which might vary between languages or even just in the same language according to different stylistical factors). To explore the interaction with the linear word order we make use of the concept of (non) projectivity instead (see §3.2).

The root-barycentre distance Δ also subsumes the DSM (“dishomogeneity”) index described by Tavoni and Chersoni (2013, §5) in quantifying the imbalance represented by “heavier” (i. e. longer and with more nodes) branches of the dependency tree. The DSM is again based on the problematic definition of geometric “left” and “right” sides of the root. Both DSM and ASM are based on the distinction of “branches” and their lengths, but a) given the nested nature of clause subordination, it appears unclear if it is sensible to consider all subtrees of a child node of the root as different, individual branches with progressive lengths, instead of directly counting the maximal depth; and b) the expansion, i. e. the width (as opposed to the depth) of a subordination level is not considered. However, the barycentre (contrary e. g. to the simple eccentricity-based centre (West, 2001, §2.1.12)) is already sensitive to the distribution of nodes at given depths, as seen from the computation of total distance. So, the barycentre can be thought of as a typologically universal generalisation of DSM and ASM, with the benefit of being informative at the same time of both to depth and width of a dependency tree.

3.2. Non Projective Gaps and Nodes

There are many equivalent definitions of (non) projectivity (Havelka, 2007, §2.1). In the following, we are interested in the notion of **gaps of non projective edges** (Havelka, 2007, §2.2): in a dependency tree $G = (N, E)$, a node $n \in N$ lies in such a gap if, for any couple of nodes i and j surrounding it (i. e. $i < n < j$ in the linear ordering) and connected by an edge (i. e. (i, j) or $(j, i) \in E$), it does not belong to the subtree rooted in the head of that edge (i or j). It is these nodes that we call non projective, and on which we base our statistics in §4, while we do not consider as such the ends of the non projective edge. Linguistically speaking, non projective nodes in our sense are “dis-

placed” words, in that they violate the expected contiguity of syntactic phrases, e. g. *eius semper populum defensantes* ‘ever defending her people’ (in *Epi-12*), where *semper* ‘ever’ intervenes inside the noun phrase *eius populum* ‘her people’.

3.3. Computational Setting

To determine Δ on a dependency tree, we have first to take its undirected representation. These and other graph operations are implemented by means of the *NetworkX* module for Python¹⁸ (Hagberg et al., 2008). We then consider only effective word nodes¹⁹ and discard punctuation marks. Further, nodes in a “horizontal” relation (*conj*, *fixed*, *flat*, *parataxis*)²⁰ are collapsed onto one single node, to account for the fact that nodes in such a relation form a block wherein syntactic distances are null, and all must have the same distance from any other node in the tree.

We base our computation of non projective nodes (§3.2) on our own Python implementation of Algorithm 1 in (Havelka, 2007, p. 26). Incidentally, we note that this computation, and thus the occurrence of non projective dependency trees, is very sensitive (more so than Δ) to the chosen annotation formalism: the same sequence of nodes might or might not yield a non projective gap according to which node is selected as the head of a phrase, and this is seen e. g. in choices like considering a copula as the head of a copular construction, or not (as usual in UD).

4. Result Description

The following charts show the root-barycentre distance Δ in relation to the sentence depth in Thomas Aquinas’ *Summa Contra Gentiles* (from the IT-TB; Figure 2) and in Dante’s Latin works (from UDante; Figure 3).

The *Summa Contra Gentiles* is an example of extensive and high-quality data, and we use it as a reference in order to put the application of the method proposed in this paper into perspective: in both Figure 2 and Figure 3, the charts show a similar increase in Δ ’s value in proportion to sentence depth. As for Dante’s Latin syntax, it can be observed that the majority of sentences has a depth range between 2 and 6 with a directly proportional Δ value between 0 and 2 with very few cases of $\Delta = 3$. This seems to be part of the various natural language phenomena of the family of “Zipf’s laws”,²¹ and should be investigated further.

The same can be observed in each one of Dante’s Latin works individually, with very slight differences

¹⁸<https://www.python.org/>

¹⁹Refer to <https://universaldependencies.org/format.html#words-tokens-and-empty-nodes-for-technicalities>.

²⁰We point to <https://universaldependencies.org/u/dep/index.html> and also refer to (Osborne, 2019, §10.3).

²¹For a general reference, see (Manning and Schütze, 1999, §1.4.3).

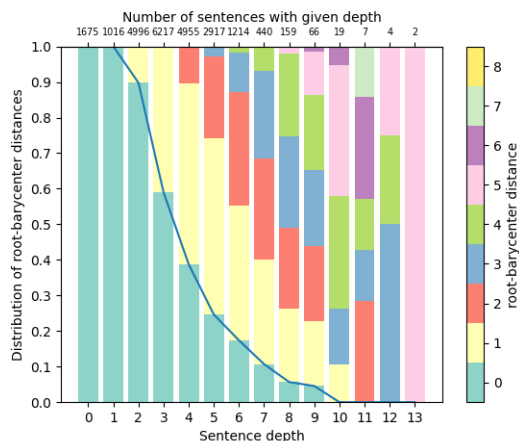


Figure 2: Δ chart of IT-TB, *Summa Contra Gentiles* (ittb-scg).

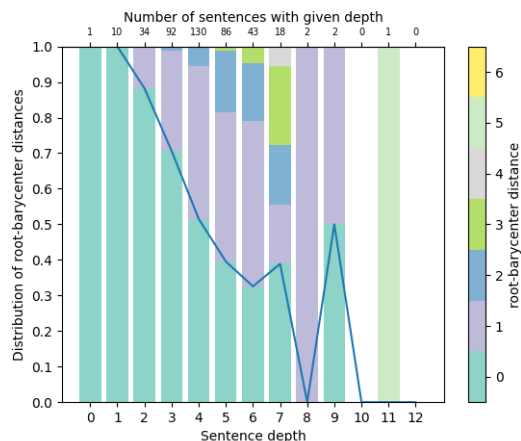


Figure 4: Δ chart of UDante, *De vulgari eloquentia* (DVE).

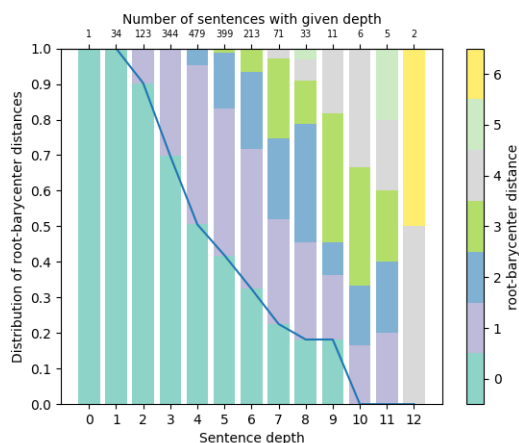


Figure 3: Δ chart over all UDante.

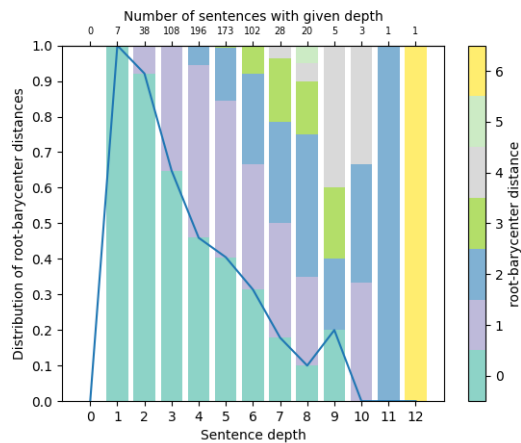


Figure 5: Δ chart of UDante, *Monarchia* (Mon).

which can nevertheless be brought into comparison. As shown in Figure 4, the *DVE* appears to be the most syntactically balanced work, even though it has the majority of non projective nodes, as defined in §3.2.

On the other hand, the prose of the *Mon*, in proportion, reaches higher values of Δ . This is shown in Figure 5, where the blue line highlighting the sentences with $\Delta = 0$ has a slightly steeper slope than that in Figure 4, and from the number of sentences where Δ 's value is 2, 3 and above.

The difference in syntactic balance is highlighted considering two of the most complex sentences in *DVE* (depth 9, $\Delta = 0$) and *Mon* (depth 12, $\Delta = 6$). Although the two sentences have a remarkably above-average depth (and the same number of tokens, 41), their dependency trees show respectively a well balanced structure in *DVE* I vi 1, and a sentence whose branches plunge to the right in a sequence of relative clauses in *Mon* II iii 16; relative clauses (a subtype of adnominal clauses) are by definition one of the types

of subordinate clauses which contribute the most to the expansion of a complex sentence, and their frequency is approximately the same in the *DVE* (2.7% of all dependency relations) and in the *Mon* (2.6%).

A similar trend is visible in the Δ chart of Dante's *Epistles* (Figure 6) and *Questio* (Figure 7), despite a significant decrease in the number of sentences, and an ensuing sparsity in the data.

The *Eclogues* (Figure 8) are the only example of Dante's Latin poetry and, although their trend is comparable to that observed for the works in prose, their syntactic depths, which are relatively low (and so their Δ value), clearly depend on the limits imposed by the verse.

5. Conclusions and future perspectives

In order to highlight the characteristics of Dante's Latin syntax, a more extensive comparison with Classical and Medieval treebanks is certainly to be called for. However, due to the lack of such resources at

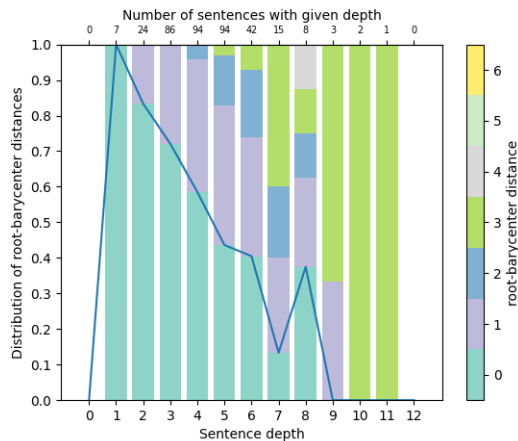


Figure 6: Δ chart of UDante, *Epistles* (Epi).

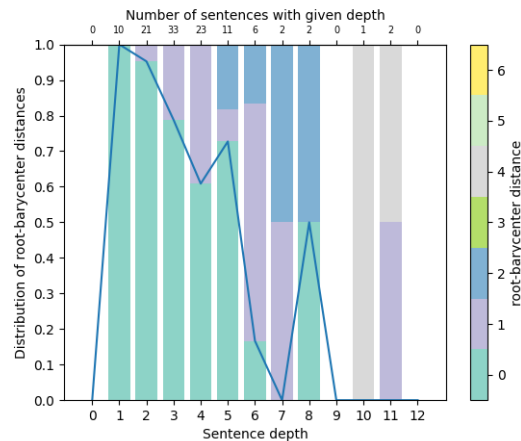


Figure 8: Δ chart of UDante, *Eclogues* (Eg).

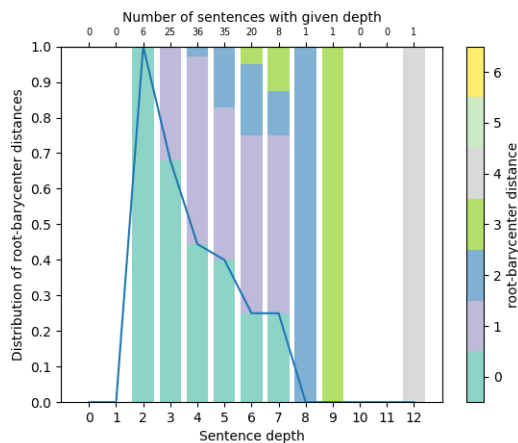


Figure 7: Δ chart of UDante, *Questio* (Que).

the current state of the art, especially within the UD framework, the method presented in this paper can be, at present, only reasonably applied to two resources: UDante and the IT-TB.

With regard to Dante, the starting point of the analysis is the study on the *suprema constructio* by Tavoni and Chersoni (2013). Even though its definition is still unclear (and will be until a treebank of all the Latin authors whom Dante quotes as examples of good Latin syntax is available), this paper translates Tavoni and Chersoni’s indices of ‘dishomogeneity’ and ‘asymmetry’ (see §3.1) into a topological model based on the mathematical structure of dependency trees. This is achieved using the notions of barycentre and depth (§3.1) and projectivity (§3.2), and the application of this model to Dante’s Latin syntax shows that it is entirely possible that Dante used the *suprema constructio* as an example of well balanced structure in his Latin writings.

Although still to be discussed within the frame of more general tendencies due to the nature of language which,

in this case, is Latin (see §4), this is a robust and reproducible corpus-based method which allows to compare the development of syntactic balance in different works and in different authors, grounding the various stylistic interpretations to a computational approach.

6. Bibliographical References

- Gabriella Albanese, et al., editors. (2019-). *Vocabolario Dantesco Latino*. Ongoing online publication at <http://www.vocabolariodantescolatino.it>.
- Alighieri, D. (1938). *De vulgari eloquentia. Ridotto a miglior lezione e commentato da Aristide Marigo, con introduzione, analisi metrica della canzone, studio della lingua e glossario*, volume VI of *Opere di Dante*. Le Monnier, Florence, Italy.
- Alighieri, D. (2011). *De vulgari eloquentia*, volume I of *Opere*. Edizione diretta da Marco Santagata. Mondadori, Milan, Italy.
- Bamman, D. and Crane, G., (2011). *The ancient Greek and Latin dependency treebanks*, pages 79–98. Theory and Applications of Natural Language Processing. Springer, Berlin/Heidelberg, Germany.
- Bamman, D., Crane, G., Passarotti, M., and Raynaud, S. (2007). Guidelines for the Syntactic Annotation of Latin Treebanks (v.1.3). Tufts Published Scholarship. Tufts University’s Digital Collections and Archives, Medford, MA, USA, February. Retrieval at <http://hdl.handle.net/10427/42683>.
- Basile, B. and Brugnoli, G. (1971). Latino. In Umberto Bosco, editor, *Enciclopedia Dantesca*, volume 3, Rome, Italy. Istituto della Enciclopedia Italiana. Accessible online at [https://www.treccani.it/enciclopedia/latino_\(Enciclopedia-Dantesca\)/](https://www.treccani.it/enciclopedia/latino_(Enciclopedia-Dantesca)/).
- Brugnoli, G. (1965). Il latino di Dante. In Casa di Dante, editor, *Dante e Roma. Atti del Convegno di studi, Roma 8-9-10 aprile 1965*, pages 51–71, Flo-

- rence, Italy. Comitato nazionale per le celebrazioni del VII centenario della nascita di Dante, Le Monnier.
- Cecchini, F. M., Korkiakangas, T., and Passarotti, M. (2020a). A New Latin Treebank for Universal Dependencies: Charters between Ancient Latin and Romance Languages. In Nicoletta Calzolari, et al., editors, *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*, pages 933–942, Marseille, France, May. European Language Resources Association (ELRA).
- Cecchini, F. M., Sprugnoli, R., Moretti, G., and Passarotti, M. (2020b). UDante: First Steps Towards the Universal Dependencies Treebank of Dante’s Latin Works. In Johanna Monti, et al., editors, *Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020, Bologna, Italy, March 1–3 2021)*, pages 99–105, Turin, Italy. Associazione italiana di linguistica computazionale (AILC), Accademia University Press. Retrievable at http://ceur-ws.org/Vol-2769/paper_14.pdf.
- Corbara, S., Moreo, A., Sebastiani, F., and Tavoni, M. (2020). *L’Epistola a Cangrande* al vaglio della *Computational Authorship Verification*: risultati preliminari (con una postilla sulla cosiddetta «XIV Epistola di Dante Alighieri»). In Alberto Casadei, et al., editors, *Nuove inchieste sull’epistola a Cangrande : atti della giornata di studi, Pisa 18 dicembre 2018*, number 2 in *ILLA – Nuove Ricerche umanistiche*, pages 153–187, Pisa, Italy. Pisa University Press. Retrievable at <http://nmis.isti.cnr.it/sebastiani/Publications/Cangrande2020.pdf>.
- Curtius, E. R. (1948). *Europäische Literatur und lateinisches Mittelalter*. A. Francke Verlag, Bern, Switzerland.
- de Marneffe, M.-C., Manning, C. D., Nivre, J., and Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, 47(2):255–308, July. Retrievable at <https://direct.mit.edu/coli/article/47/2/255/98516/Universal-Dependencies>.
- De Mauro, T. and Allasia, C. (2005). *La fabbrica delle parole*. UTET, Turin, Italy.
- Eckhoff, H. M., Bech, K., Bouma, G., Eide, K., Haug, D., Haugen, O. E., and Jøhndal, M. (2018). The PROIEL treebank family: a standard for early attestations of Indo-European languages. *Language Resources and Evaluation*, 52(1):29–65.
- Fioravanti, G. (2017). Alberto di Sassonia, Biagio Pelacani e la *Questio de aqua et terra*. *Studi Danteschi*, LXXXII:81–97.
- Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux, et al., editors, *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA, USA. Retrievable at http://conference.scipy.org/proceedings/SciPy2008/paper_2/. Software and documentation at <https://networkx.org/>.
- Havelka, J. (2007). *Mathematical Properties of Dependency Trees and their Application to Natural Language Syntax*. Ph.D. thesis, Univerzita Karlova – Matematicko-fyzikální fakulta, Prague, Czech Republic, June. Retrievable at <https://dspace.cuni.cz/handle/20.500.11956/12614?locale-attribute=en>.
- Koschützki, D., Lehmann, K. A., Peeters, L., Richter, S., Tenfelde-Podehl, D., and Zlotowski, O. (2005). Centrality Indices. In Ulrik Brandes et al., editors, *Network Analysis: Methodological Foundations*, number 3418 in *Lecture Notes in Computer Science*, pages 16–61, Berlin/Heidelberg, Germany. Springer.
- Ledgeway, A. (2012). *From Latin to Romance*. Number 1 in *Oxford studies in historical and diachronic linguistics*. Oxford University Press, Oxford, UK.
- Manning, C. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA, May. Companion website: <https://nlp.stanford.edu/fsnlp/>.
- Nilsson, J. and Nivre, J. (2008). MaltEval: an Evaluation and Visualization Tool for Dependency Parsing. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA). Retrievable at http://www.lrec-conf.org/proceedings/lrec2008/pdf/52_paper.pdf.
- Osborne, T. (2019). *A Dependency Grammar of English*. John Benjamins, Amsterdam, Netherlands; Philadelphia, PA, USA.
- Paratore, E. (1965). Il latino di Dante. *Cultura e scuola*, IV:94–124.
- Passarotti, M., Pedonese, G., and Sprugnoli, R. (2021). Le opere di Dante tra annotazione linguistica e web semantico. *Linguistica e letteratura*, XLVI:45–71.
- Passarotti, M., (2019). *The Project of the Index Thomisticus Treebank*, pages 299–320. Number 10 in *Age of Access? Grundfragen der Informationsgesellschaft*. De Gruyter Saur, Berlin, Germany; Boston, MA, USA. Retrievable at <https://www.degruyter.com/view/book/9783110599572/10.1515/9783110599572-017.xml>.
- Ponti, E. M. and Passarotti, M. (2016). Differentia compositionem facit. A slower-paced and reliable parser for Latin. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 683–688, Portorož, Slovenia, May. European Language Resources Association (ELRA).

- Straka, M., Hajič, J., and Straková, J. (2016). UD-Pipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4290–4297, Portorož, Slovenia, May. European Language Resources Association (ELRA).
- Tavoni, M. and Chersoni, E. (2013). Ipotesi d'interpretazione della «suprema constructio» (*De vulgari eloquentia* II vi). *Studi di grammatica italiana*, XXXI–XXXII:131–158.
- Tavoni, M. (2011). DanteSearch: il corpus delle opere volgari e latine di Dante lemmatizzate con marcatura grammaticale e sintattica. In Anna Cerbo, et al., editors, *Lectura Dantis 2002-2009. Omaggio a Vincenzo Placella per i suoi settanta anni*, volume 2 (2004–2005), pages 583–608, Naples, Italy. Il Torcoliere – Officine Grafico-Editoriali di Ateneo.
- West, D. B. (2001). *Introduction to Graph Theory*. Prentice Hall, Englewood Cliffs, NJ, USA, second edition. Companion site at <https://faculty.math.illinois.edu/~west/igt/>.