The Rise and Fall of Dependency Parsing in Dante Alighieri's *Divine Comedy*

Claudia Corbetta[1], Marco Passarotti[2], Giovanni Moretti[2]
Università di Pavia/Bergamo[1], Università Cattolica del Sacro Cuore[2],
claudia.corbetta@unibg.it {marco.passarotti, giovanni.moretti}@unicatt.it

Abstract

In this paper, we conduct parsing experiments on Dante Alighieri's *Divine Comedy*, an Old Italian poem composed between 1306-1321 and organized into three *Cantiche —Inferno*, *Purgatorio*, and *Paradiso*. We perform parsing on subsets of the poem using both a Modern Italian training set and sections of the *Divine Comedy* itself to evaluate under which scenarios parsers achieve higher scores. We find that employing in-domain training data supports better results, leading to an increase of approximately +17% in Unlabeled Attachment Score (UAS) and +25-30% in Labeled Attachment Score (LAS). Subsequently, we provide brief commentary on the differences in scores achieved among subsections of *Cantiche*, and we conduct experimental parsing on a text from the same period and style as the *Divine Comedy*.

Keywords: Parsing, Dependency syntax, Old Italian, Modern Italian, *Divine Comedy*

1. Introduction

The *Divine Comedy*[1], an Old Italian[2] poem authored by Dante Alighieri, was composed in the period between 1306 and 1321. This seminal work comprises three *Cantiche*: *Inferno*, *Purgatorio*, and *Paradiso*. Each *Cantica* is subdivided into *Canti*, culminating in a total of 100 (34 in *Inferno*, 33 in *Purgatorio*, and 33 in *Paradiso*)[3]. Recognized as a foundational pillar of Italian literature, the language of the *Divine Comedy* plays a pivotal role in the evolution of the Italian language.

A linguistic annotation of the *Divine Comedy* is provided by DanteSearch (Tavoni, 2011), an online corpus[4] containing all the works of Dante Alighieri. DanteSearch employs a tagset to identify parts of speech (PoS) and morphological features of words[5] and provides a clause-based syntactic annotation style, wherein the functions of clauses within the sentence (e.g., declarative, interrogative, exclamative) are recorded[6].

Nevertheless, the annotation schema and tagset of DanteSearch are not fully compatible with other styles, such as the one used in the Universal Dependencies initiative[7] (UD), which is currently the standard de facto schema for syntactically annotated corpora. UD is an annotation framework designed to establish a universal formalism for dependency-based syntactic annotation (De Marneffe et al., 2021). Its primary objective is to facilitate cross-linguistic comparison, starting by collecting linguistic information into a treebank, a linguistically annotated corpus containing several layers of annotation such as lemmatization, PoS and (dependency) syntax annotation.

In the UD collection, the first and sole treebank documenting Old Italian is the *Divine Comedy*. Specifically, this treebank, referred to as Italian-Old in UD, encompasses the first *Cantica* of the *Divine Comedy*, namely *Inferno*. The creation of the treebank for the *Divine Comedy* (Corbetta et al., 2023) leveraged pre-existing annotated data from DanteSearch. While PoS and lemmas were semi-automatically converted from DanteSearch to the UD format, the dependency-based syntactic annotation was conducted anew.

Besides the need to change the syntactic annotation style from clause level to word level[8], the UD-like annotation of *Inferno* was conducted fully

[1]This paper is the result of the collaboration between the three authors. For academic purposes, Claudia Corbetta is responsible of Sections 2,3,4; Marco Passarotti of Sections 1,5; Giovanni Moretti developed the tri-gram and sub-tree extraction script and built the Stanza Model of *Inferno* IV-XXXIV.

[2]In this paper, the language of the *Divine Comedy* is referred to as Old Italian. For an in-depth understanding of the language of the *Divine Comedy*, see (Manni, 2013).

[3]Refer to (Inglese, 2012) for an introductory overview of the poem.

[4]`https://dantesearch.dantenetwork.it`

[5]To gain a deeper understanding of the concept of "word" as attested in DanteSearch, we refer to (Tavoni, 2011).

[6]For a comprehensive understanding of the clause-based annotation scheme, please see (Gigli, 2004).

[7]`https://universaldependencies.org.`

[8]As previously mentioned, the clause-based syntactic annotation style utilized by DanteSearch is not compatible with that of UD, which is word-based. For a more in-depth understanding of the distinction, please see (Corbetta et al., 2023).

LT4HALA 2024@LREC-COLING-2024, pages 50–56
25 May, 2024.

manually for two main reasons: (i) to enhance the annotators' skills through steady confrontation with data; and (ii) to prevent biases in the annotation work that could arise from using a pre-parsed text. Specifically, we did not use the trained models of parsers developed from the UD treebanks for Modern Italian.

Having completed the manual annotation for *Inferno*, this paper evaluates the performance of models trained on Modern Italian treebanks available in UD, as well as models trained on subsets of *Inferno* itself. This evaluation aims to ascertain whether one approach is preferred over the other for assisting in the annotation of the remaining parts of the *Divine Comedy*, specifically *Paradiso*[9]. Additionally, in the context of future work, we aim to explore whether using a parser based on the *Divine Comedy* could be beneficial for annotating similar texts from the same period.

The paper is organized as follows. Section 2 describes tests of parsing on *Inferno* with Modern Italian data. Section 3 describes how we selected the subset upon which we conduct parser experiments and illustrates how we calculated the correlation degree among the subset and their respective *Cantica*. Section 4 reports the results of experimenting parsing respectively with models trained on the *Divine Comedy* data and with models trained on Modern Italian data. We compare scores among the *Cantiche* and conduct parsing tests on a poem from the same period. The final section 5 summarizes the results and highlights future directions of research.

2. Parsing *Divine Comedy* Text with Modern Italian Data: a Journey through *Inferno*

The comparison between Old Italian and Modern Italian, particularly concerning syntax, has been a topic of debate[10]. The examination of potential distinctions between Old Italian and Modern Italian language lies beyond the scope of this paper. Our current investigation focuses on evaluating the syntactic accuracy of models trained on Modern Italian data for parsing *Inferno*.

While in UD the sole treebank containing Old Italian data is Italian-Old, consisting of *Inferno*, Modern Italian is covered by multiple treebanks, representing diverse styles and genres[11]. We specify that among all Modern Italian treebanks, none represents the same genre as *Divine Comedy*, namely the poetic genre, which might affect negatively the accuracy rates of the trained models.

As *Inferno* is the sole manually annotated treebank of Old Italian available, we test the accuracy of parsers using a training set based on Modern Italian data. We parse *Inferno* using two different parsers. We employ UDPipe1 (Straka et al., 2016; Straka and Straková, 2017), which is a trainable pipeline for tokenization, tagging, lemmatization and dependency parsing of CoNLLU-files[12], and Stanza (Qi et al., 2020), a neural network pipeline, that includes, among other functionalities, tokenization, tagging, lemmatization and dependency parsing[13]. For both UDPipe1 and Stanza, we only perform parsing, retaining the tokenization, lemmas, PoS and morphological features of the manually annotated text. We build models using UDPipe1 and Stanza based on training sets provided by three major Modern Italian treebanks (six models in total): ISDT (Bosco et al., 2013), VIT (Tonelli et al., 2008) and Par-TUT(Bosco et al., 2012)[14]. We evaluate the performance of the two parsers, by averaging the accuracy rates of their trained models (two evaluation rates in total). To evaluate the accuracy of the output, we rely on `eval.py`[15].

Table 1 reports the scores.

Inferno	UDPipe1	Stanza
UAS	65.28	65.16
LAS	56.98	50.85

Table 1: Accuracy metrics of *Inferno* with UDPipe1 and Stanza.

Considering that the average UAS (Unlabeled Attachment Scores) and LAS (Labeled Attachment

[9]We completed the annotation of *Purgatorio* and it is scheduled for publication in the upcoming next release of UD. See https://universaldependencies.org/release_checklist.html.

[10]We refer to the Preface of (Salvi and Renzi, 2010) for an introduction to Old Italian and its differences with Modern Italian. For an overview of syntactic peculiarities of Old Italian syntax, we refer to (Dardano and Frenguelli, 2002).

[11]For detailed information about Modern Italian treebanks in UD, see https://universaldependencies.org/it/index.html.

[12]https://github.com/ufal/udpipe.

[13]https://stanfordnlp.github.io/stanza/index.html.

[14]Refer to https://github.com/UniversalDependencies/UD_Italian-ISDT for ISDT; https://github.com/UniversalDependencies/UD_Italian-VIT for VIT and https://github.com/UniversalDependencies/UD_Italian-parTUT for Par-TUT.

[15]The `eval.py` is designed to assess the accuracy of a UD tokenizer, lemmatizer, tagger and parser against a gold-standard data. The script is available at https://github.com/UniversalDependencies/tools/blob/master/eval.py.

Scores)[16] are around 65.22 and 53.91 respectively, we have decided to challenge the results for Modern Italian by attempting to increase the scores. To do so, we utilize samples from the *Divine Comedy* as training set.

3. Data: Evaluating Correlation Degree

We select a subset of three *Canti* as test set, comprising 9% of the respective *Cantiche*[17], which we demonstrate to be adequately representative of their respective *Cantica*.

In order to evaluate the correlation degree of each subsection with its respective *Cantica*, we examine tri-gram variation in PoS tagging and subtree label attachment. The evaluation is performed by calculating the Pearson correlation coefficient for each measure, comparing the linguistic features of the subset with those of its corresponding *Cantica*. This approach provides a quantitative measure of how closely the linguistic characteristics align between the subsection and the complete *Cantica*.

3.1. Part of Speech Tri-gram Detection

We assess the degree of correlation for tri-gram PoS by converting DanteSearch tagset into UD PoS. For this task, a direct automated conversion from DanteSearch to UD PoS is applied. This means that the conversion was performed without considering the different criteria of PoS assignment between DanteSearch and UD. For instance, we do not differentiate cases such as possessive adjectives, which are tagged as adjectives in DanteSearch but classified as determiners in UD[18].

More specifically, the tri-grams analysis of PoS is conducted on the subset of *Canti* I-III of *Inferno*, *Purgatorio* and *Paradiso*, corresponding to the aforementioned 9% of the *Cantiche*. This analysis is then compared with the tri-gram distribution of the respective *Cantica*.

PoS tri-grams are extracted at sentence level, using full stops for sentence splitting[19]. For instance, Table 2 reports PoS tri-grams for the following sentence.

> *Se' savio; intendi me' ch'i' non ragiono.*
> (*Inf*. II, v. 36)
>
> You're wise; you know far more than what I say.

Tri-gram of words	PoS tri-gram
Se'/savio/intendi	AUX/ADJ/VERB
savio/intendi/me'	ADJ/VERB/ADV
intendi/me'/ch'	VERB/ADV/SCONJ
me'/ch'/i'	ADV/SCONJ/PRON
ch'/i'/non	SCONJ/PRON/ADV
i'/non/ragiono	PRON/ADV/VERB

Table 2: The extraction of tri-grams from a sentence in *Inferno*.

Tri-grams of each subsection are then listed according to their frequency and compared with the tri-gram rankings of the respective *Cantica*, evaluating the Pearson correlation coefficient (Brezina, 2018) to estimate their correlation degree. To mitigate data sparsity due to the different size of the texts compared, we exclude the tri-grams belonging to the less frequent 5% of the total[20].

Table 3 reports the Pearson correlation of each subset in respect with its *Cantica*.

Inferno	Purgatorio	Paradiso
0.835	0.845	0.868

Table 3: Pearson correlation for tri-gram PoS.

As Pearson coefficient is > 0.5 (Brezina, 2018, p. 144), we can consider the correlation to be strong and generalize that the PoS tri-gram distribution of each subset of *Canti* I-III correlates with its respective *Cantica*.

3.2. Sub-tree Label Attachment

We also assess the correlation degree by examining the syntactic structure, specifically sub-tree dependency relations, in the subsection I-III of *Inferno* and I-III *Purgatorio* compared with the corresponding *Cantica*. We abstain from conducting correlation for *Paradiso* since its syntax is presently under development. We assume that the sub-tree label correlation evaluated within *Inferno* and *Purgatorio* could be consistent with the other *Cantica*, in agreement with the results shown in the PoS correlation.

[16] Refer to (Buchholz and Marsi, 2006) for an insight into syntactic metrics.

[17] The number of tokens in each subset (I-III) is 3561 tokens for *Inferno*, 3622 for *Purgatorio*, and 3484 for *Paradiso*.

[18] The described procedure will not have a negative impact on the evaluation, as we maintain a unified PoS tagging system, specifically the one adopted by DanteSearch, and we consistently employ such scheme to analyze the tri-gram correlation within the first three *Canti*.

[19] This means that the PoS of the last word of a sentence is the final item of a tri-gram, while the PoS of the first word of a sentence serves as the first item of a tri-gram.

[20] This results in excluding around the 1700 tri-grams out of the total of approximately 32300. More precisely, we exclude 1726 out of 32564 for *Inferno*; 1729 out of 32428 for *Purgatorio* and 1701 out of 32027 for *Paradiso*.

When referring to sub-tree dependency relations, we denote a sub-tree composed of the PoS of a governor node (n1), the dependency relation[21] of the node n1 with its dependent (deprel, such as `nsubj` for the subject relation), and the PoS of the dependent node (n2), following the schema:

ragiono –> `nsubj` –> i'

VERB –> `nsubj` –> PRON

More precisely, the extraction of sub-tree labels is performed for each syntactic node (except for punctuation, marked with the deprel `punct`). For each node, a triple is extracted, consisting of the node (n1), a dependent node (n2) and their dependency relation. Subsequently, we derive the PoS of the involved nodes.

Following the approach used for tri-grams, we subsequently apply Pearson correlation to assess the correlation degree of sub-tree labels between the two. Similarly to the PoS tri-gram, we exclude sub-trees that belong to the least frequent 5% of the total[22]. Pearson correlation showed in Table 4 is > 0.5, namely 0.744 for *Inferno* and 0.737 for *Purgatorio*, highlighting a strong correlation between the sub-tree dependency labels of the first three *Canti* of *Inferno* and *Purgatorio* and their entire *Cantica*.

Inferno	Purgatorio
0.772	0.794

Table 4: Pearson correlation for sub-tree labels.

Given the high Pearson coefficient observed in both tri-gram correlation and sub-tree labels (limited to *Inferno* and *Purgatorio*), we conclude that the first three *Canti* might be partially considered representative of the respective *Cantica*.

4. Parsing and Evaluation: Examining the First Three *Canti*

Given the high correlation degree between the first three *Canti* and their respective *Cantiche*, we proceed with parsing experiments and evaluation metric checks to see whether, by using subsets of the *Divine Comedy* as training data, we can improve the UAS and LAS scores obtained with Modern Italian training data and reported in Table 1.

4.1. *Divine Comedy* on *Divine Comedy*

We train both UDPipe1 (UDP) and Stanza (Stan) on a training set consisting of *Canti* IV-XXXIV of *Inferno*, encompassing the 30% of the all *Divine Comedy* and 91% of *Inferno*[23]. Subsequently, we test the two models on the first three *Canti* of each *Cantica*, namely *Inferno* I-III (Inf), *Purgatorio* I-III (Purg) and *Paradiso* I-III (Par) and evaluate the syntax metrics, namely LAS and UAS, with respect to the gold standard of each test set[24]. The evaluation is performed using `eval.py` for the output of UDPipe1 model. In the case of Stanza, the evaluation is executed automatically after each training run[25].

Table 5 shows LAS and UAS of each subset, i.e., I-III of *Inferno* (Inf), *Purgatorio* (Purg) and *Paradiso* (Par), for both UDPipe1 (UDP) and Stanza (Stan) model.

It is noteworthy that the scores provided by both UDPipe1 and Stanza in Table 5 are significantly higher when compared with the scores obtain from model trained on Modern Italian data on *Inferno* (see Table 1). The boost of the Stanza model trained on the *Divine Comedy* data is 19.81 for UAS and 29.21 for LAS[26] compared to the Stanza model trained on Modern Italian data. Regarding UDPipe1, we observe an increase of 14.35 scores for UAS and 16.56 scores for LAS in favor of models trained on *Divine Comedy*[27].

We replicate the test using only the Stanza model with the same training set of Modern Italian used for the data in 2, testing it on the subsets of *Inferno* I-III, *Purgatorio* I-III and *Paradiso* I-III.

As shown in Table 6, scores obtained with the training set of Modern Italian on the subsets *Inf*, *Purg*, and *Par*, reflect the scores obtained for the parsing of *All Inferno*, reported in Table 1. This confirms that using part of the text as the training set yields better results than using Modern Italian data.

It is also interesting to note that the scores across the *Cantiche* flow both in Table 5 and Table 6, being higher for *Inferno*, followed by *Paradiso*, and then *Purgatorio*. We briefly comment fluctuations in Subsection 4.2.

[21] A list of dependency relations and the specific meaning of each label is documented in UD.

[22] This implies that we do not consider 1764 sub-trees out of 33387.

[23] The training set consists of 1118 sentences and 37806 syntactic words.

[24] The gold standards of *Purgatorio* I-III and *Paradiso* I-III were manually annotated by an annotator with competence in Old Italian.

[25] For detailed information on the evaluation in Stanza, please see https://stanfordnlp.github.io/stanza/training_and_evaluation.html#evaluation.

[26] We considered the average of both LAS and UAS scores for *Inf*, *Purg* and *Par* subsets in Table 5, precisely 84.97 for UAS and 80.06 for LAS.

[27] The average of LAS and UAS scores for the subsets *Inf*, *Purg*, and *Par* are respectively 73.54 and 79.63.

	Inf		Purg		Par	
Metr.	UDP	Stan	UDP	Stan	UDP	Stan
UAS	82.65	87.73	77.93	82.67	78.50	84.50
LAS	77.87	84.02	71.42	77.31	71.33	78.85

Table 5: LAS and UAS scores of each subset parsed with UDPipe1 and Stanza.

	Inf	Purg	Par
UAS	69.05	66.28	67.74
LAS	56.14	53.30	54.31

Table 6: LAS and UAS scores of each subset parsed with Stanza model trained on Modern Italian.

4.2. Comparing Metrics across *Cantiche*

By analyzing syntactic metrics across *Cantiche*, we notice that scores flow throughout the samples of *Inferno*, *Purgatorio*, and *Paradiso*. Such fluctuations are evident in both datasets parsed with a training dataset composed of a section of *Inferno* (Table 5) and the one trained with Modern Italian data (Table 6).

In this Section, we briefly comment on the data in Table 5, namely on metrics achieved from models trained on *Divine Comedy* data[28]. The metrics presented in Table 5 demonstrate an enhanced performance under an "in-domain" condition, specifically when the training and test sets pertain to the same *Cantica*, *Inferno*. When comparing the UAS and LAS scores of *Inferno* with those of *Purgatorio* and *Paradiso*, *Inferno*'s metrics show a boost of 4.14 (Stanza) and 4.44 (UDPipe1) in LAS, and of 5.94 (Stanza) and 6.50 (UDPipe1) scores in UAS[29].

Examining closely the differences among *Purgatorio* and *Paradiso*, we also observe that *Paradiso* outperforms *Purgatorio*. Specifically, for both UDPipe1 and Stanza models, *Paradiso* experiences an improvement of 0.57 and 1.83 in UAS, respectively. The LAS score boost achieved by the Stanza model supports the observed trend in UAS metrics, with *Paradiso* LAS achieving a superior score of 1.54 points compared to *Purgatorio*. Contrary to the trend, the UDPipe1 LAS score seems to exhibit a slightly better performance in *Purgatorio* than in *Paradiso*, but the difference of 0.09 in score is very low.

The data presented suggest that syntactic structures of *Paradiso* seem to be more akin to *Inferno* than *Purgatorio* is to *Inferno*, especially for the first three *Canti* of the *Cantica*. However, such a claim deserves to be substantiated through additional studies.

4.3. Experimenting outside the *Divine Comedy*: Testing Guido Cavalcanti's Poem

To verify the efficiency of the Stanza model trained on the *Divine Comedy* data, we test it on a text from the same period and style as Dante Alighieri's poem. We select a text by Guido Cavalcanti (1259-1300), a poet contemporary to Dante and belonging to the same socio-cultural milieu[30]. The selected text is "Voi che per li occhi mi passaste il core", a poem in Old Italian, specifically Old Florentine, consisting of 111 syntactic words.

We parse the poem with Stanza model trained on all Inferno and with Stanza models trained on different Modern Italian treebanks[31] and we evaluate the syntactic metrics. Tokenization, lemmatization, PoS tagging, and morphological features are provided to the model, which is solely tasked with performing syntactic tasks.

	Stan All Inf	Stan Mod It
UAS	86.49	66.37
LAS	75.68	48.65

Table 7: Metrics in Cavalcanti's poem with Stanza model trained on All *Inferno* and Stanza model trained on Modern Italian data.

As shown in Table 7, Stanza model trained on All Inferno performs better than Modern Italian one. The boost is significantly around 20.12 for UAS and 27.03 for LAS.

Despite the small sample size, the boost is promising. We will further investigate and experiment by testing on larger samples and expanding the domain to include more authors and texts of the same period to understand whether the *Divine Comedy* might be representative enough.

[28] Discussion of metrics achieved from Modern Italian data will be left for further studies.

[29] To calculate the boost of *Inferno*'s scores, we consider an average among the UAS and the LAS scores of *Purg* and *Par* scores.

[30] We refer to (Cavalcanti, 2011) for an introduction of Guido Cavalcanti and his rhymes.

[31] After parsing the poem with models trained on respectively ISDT, VIT, Par-TUT, we calculate an average of the scores of all Modern Italian models.

5. Conclusion

In this paper, we parse sections of the *Divine Comedy*, comparing the accuracy of models trained on Modern Italian data with those trained on portions of the *Divine Comedy* itself.

Firstly, our findings reveal that employing parsers trained on texts from the *Divine Comedy*, namely within their respective domain, result in higher accuracy. Such trend confirms the literature stating that having in-domain training data facilitates parsing results (Khan et al., 2013b,a), particularly when dealing with texts from the same author (Mambrini and Passarotti, 2012). We can therefore conclude that, at the current state of the art, despite having a larger amount of Modern Italian treebanks, using Modern Italian training set to parse the *Divine Comedy* does not result in better parsing outcomes.

Additionally, the data obtained from the comparison among the first three *Canti* of each *Cantica* highlight a greater proximity between the syntax of the first three *Canti* of *Paradiso* and the first ones of *Inferno*, compared to *Purgatorio*. However, even though we have demonstrated the representativeness of the first three *Canti* with the respective *Cantica*, the analyzed data do not allow us to identify a specific trend sufficient to draw conclusions about the possible proximity or distance between the syntax of the all three *Cantiche*.

Lastly, we conduct a brief experiment on a text contemporaneous with the *Divine Comedy*, illustrating the superiority of utilizing a model trained on similar chronological and textual types over models of Modern Italian.

As potential future work, we will investigate whether augmenting the training data by merging datasets from both Old and Modern Italian, notwithstanding the diversity in genre, will result in enhanced parsing accuracy. Moreover, further studies, along with additional annotated data[32], are necessary to ascertain the relationship between the results and the diversity of genres. Future research endeavors will be dedicated to delving deeper into these aspects.

6. Acknowledgments

We would like to thank the anonymous reviewer for the accurate suggestions.

7. Bibliographical References

[32]To address the variety of genres across Old and Modern Italian, we need annotated data for both non-poetic Old Italian literature and Modern Italian poetry, currently unavailable.

Cristina Bosco, Simonetta Montemagni, and Maria Simi. 2013. Converting Italian treebanks: Towards an Italian Stanford dependency treebank. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 61–69, Sofia, Bulgaria. Association for Computational Linguistics.

Cristina Bosco, Manuela Sanguinetti, and Leonardo Lesmo. 2012. The parallel-TUT: a multilingual and multiformat treebank. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1932–1938, Istanbul, Turkey. European Language Resources Association (ELRA).

Vaclav Brezina. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Sabine Buchholz and Erwin Marsi. 2006. CoNLL-X Shared Task on Multilingual Dependency Parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 149–164, New York City, NJ, USA. Association for Computational Linguistics.

Guido Cavalcanti. 2011. *Rime*. A cura di Giorgio, Inglese and Roberto, Rea. Carocci; Critical Edition.

Claudia Corbetta, Marco Carlo Passarotti, Flavio Massimiliano Cecchini, and Giovanni Moretti. 2023. Highway to hell. towards a universal dependencies treebank for dante alighieri's comedy. In *Proceedings of CLiC-it 2023: 9th Italian Conference on Computational Linguistics*, Venice, Italy. Associazione Italiana di Linguistica Computazionale.

Maurizio Dardano and Gianluca Frenguelli. 2002. *SintAnt. La sintassi dell'italiano antico*. ARACNE.

Marie-Catherine De Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Sara Gigli. 2004. Codifica sintattica della commedia dantesca. *PhD diss., Università di Pisa*.

Giorgio Inglese. 2012. *Dante: guida alla Divina Commedia. Nuova edizione*. Carocci, Roma, Italy.

Mohammad Khan, Markus Dickinson, and Sandra Kübler. 2013a. Towards domain adaptation for parsing web data. In *Proceedings of the International Conference Recent Advances in Natural*

Language Processing RANLP 2013, pages 357–364, Hissar, Bulgaria. INCOMA Ltd. Shoumen, BULGARIA.

Mohammad Khan, Markus Dickinson, and Sandra Kuebler. 2013b. Does size matter? text and grammar revision for parsing social media data. In *Proceedings of the Workshop on Language Analysis in Social Media*, pages 1–10, Atlanta, Georgia. Association for Computational Linguistics.

Francesco Mambrini and Marco Carlo Passarotti. 2012. Will a parser overtake Achilles? First experiments on parsing the ancient Greek dependency treebank. In *Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories (TLT11). 30 November–1 December 2012, Lisbon, Portugal*, pages 133–144. Edições Colibri.

Paola Manni. 2013. *La lingua di Dante*. il Mulino, Bologna, Italy.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. pages 101–108.

Giampaolo Salvi and Lorenzo Renzi, editors. 2010. *Grammatica dell'italiano antico*. il Mulino, Bologna, Italy.

Milan Straka, Jan Hajič, and Jana Straková. 2016. UDPipe: 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. European Language Resources Association (ELRA).

Milan Straka and Jana Straková. 2017. Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada. Association for Computational Linguistics.

Mirko Tavoni. 2011. *DanteSearch: il corpus delle opere volgari e latine di Dante lemmatizzate con marcatura grammaticale e sintattica*, volume 2 (2004–2005), pages 583–608. Il Torcoliere – Officine Grafico-Editoriali di Ateneo, Napoli, Italy.

Sara Tonelli, Rodolfo Delmonte, and Antonella Bristot. 2008. Enriching the venice Italian treebank with dependency and grammatical relations. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).