

A Treebank of Asia Minor Greek

Eleni Vligouridou, Inessa Iliadou, Çağrı Çöltekin

University of Tübingen
Tübingen, Germany

eleni.vligouridou@student.uni-tuebingen.de, gianna_iliadou95@yahoo.gr,
ccoltekin@sfs.uni-tuebingen.de

Abstract

Asia Minor Greek (AMG) dialects are endangered dialects rich in history and culture that face a dire struggle for preservation due to declining speaker base and scarce linguistic resources. To address this need, we introduce a Universal Dependencies treebank of Phrasiot Greek, one of the severely endangered AMG dialects. The present treebank is fully manually annotated and currently consists of 350 sentences from six fairy tales in Phrasiot dialect. Besides describing the treebank and the annotation process, we provide and discuss interesting phenomena we observed in the treebank. Most phenomena we discuss are related to contact-induced linguistic changes that these dialects are well known for. Beyond linguistic inquiry, like other treebanks for truly low-resource languages, the AMG treebank we present offers potentials for diverse applications, such as language preservation and revitalization, as well as NLP tools that have to be developed with scarce resources.

Keywords: Phrasiot, Cappadocian, Asia Minor Greek, Greek, treebank, Universal Dependencies

1. Introduction

In this paper, we present a manually annotated treebank of Phrasiot, one of the Asia Minor Greek (AMG) dialects, following the Universal Dependencies (UD) framework (Nivre et al., 2016; De Marnette et al., 2021). AMG dialects were originally spoken in Asia Minor and currently most language varieties are spoken by only a limited number of speakers in Greece. Besides their endangered status, these dialects are also interesting as a historically isolated variety of Greek under strong influences of other languages, mainly Turkish, due to language contact (Janse, 2009).

Treebanks are typically created for building automated tools (e.g., morphological analyzers or parsers) for morphosyntactic analysis of a language. In case of endangered and lesser studied languages, however, they may also serve as an (incomplete) linguistic description; they can support language learning, preservation and revitalization efforts; and allow researchers to test linguistic hypotheses (e.g., related to theoretical, typological or historical aspects of the language). To support these goals, we present a manually annotated UD treebank of 350 sentences from six fairy tales. The treebank is annotated both for morphology and syntax. Although an earlier treebank is documented by Sampanis and Prokopidis (2021), to our knowledge, our treebank is the first treebank ready to be released publicly.¹

In the remainder of this paper, we give a brief

¹The current version of the treebank is publicly available at <https://github.com/iscl-lrl/amg-treebank>. The treebank will be released in the UD repository with the UD version 2.14 (May 2024).

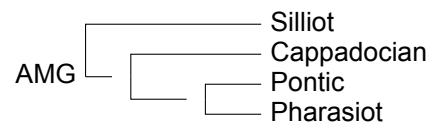


Figure 1: The classification of AMG dialects by Dawkins (1916).

description of the AMG dialects and review existing resources and descriptive work on Asia Minor Greek dialects in Section 2, Section 3 describes the treebank and the annotation process, Section 4 provides a preliminary analysis and discussion of some of the examples, highlighting the effects of language contact, and Section 5 provides a short conclusion with potential future uses and improvements of the treebank.

2. Asia Minor Greek Dialects

Asia Minor Greek dialects constitute the dialects that were spoken in (inner) Anatolia, including Cappadocian, Phrasiot, Silli, and Pontic Greek. The most comprehensive study of the AMG dialects is conducted by Dawkins (1916) based on his fieldwork in central Anatolia. The classification of these dialects by Dawkins (1916) is shown in Figure 1.²

²The classification presented here diverges from the one presented in Glottolog (Hammarström et al., 2024). Our choice of the source of this classification, at this point, is rather arbitrary. However, linguistically annotated corpora as the one presented here may also be instrumental for a more accurate classification of the AMG dialects.

The AMG dialects follow an interesting historical development since they are separated from the rest of the Greek speaking populations since the 11th and 12th century (Janse, 2009). By early 20th century, almost all speakers were bilinguals, and the dialects they spoke were strongly affected by language contact (Dawkins, 1916). After the population exchange between Greece and Turkey in 1923, AMG speakers were forced to move to Greece, adopting standard Greek in daily communication. This, in turn, led to the diminishing number of speakers, and endangered status of these language varieties. Until recently, some of these dialects were thought to be extinct. However, Mark Janse and Dimitris Papazachariou reported fluent Cappadocian speakers in Simantra in South Macedonia and Thessaly (Janse, 2009). Cappadocian Greek is spoken by approximately 10 000 speakers according to UNESCO.³ Similarly, the Phrasiot Greek is reported to have about 25 known speakers by Bagri ık (2018).

In addition to the historical interest and their endangered status, AMG dialects also show a number of remarkable linguistic properties due to long-term linguistic contact with Turkish, a language with major typological differences from Greek. As well as lexical borrowings that are abundant in our treebank, there are some phonological, morphological and syntactic phenomena observed in AMG dialects as results of linguistic contact. Some of the phonetic and phonological aspects include, elision of the letter η in the post tonic syllable (Dawkins, 1916, p.149), conversion of o to ou when an intermediate consonant is eliminated (Andriotis, 1948, p.20), substitution of the unstressed η with ε (Kontossopoulos, 1981, p.9), fusion of diphthongs (e.g., ε instead of ια, Andriotis (1948, p.18) and instances of vowel harmony (e.g.,  ερεϑή ‘Sunday’) (Revithiadou et al., 2006). Our treebank contains many examples of these phenomena.

The language contact on AMG dialects is not only limited to lexical borrowings and sound changes, but also morphosyntactic patterns (Ralli, 2020), such as traces of agglutinative morphology (Janse, 2004; Ralli, 2009; Karatsareas, 2011, 2016; Revithiadou et al., 2017; Janse, 2019), borrowing of structural elements like adpositions and changes to word order (Bagri ık, 2018; Janse, 2020; Melissaropoulou and Ralli, 2020). These phenomena can easily be quantified in a dependency treebank, indicating the value of treebanks as resources for investigating such morphosyntactic phenomena quantitatively. This particularly makes the linguistically annotated resources for the AMG dialects, a good candidate as case stud-

³<https://en.wal.unesco.org/languages/cappadocian-greek> (accessed 2023-10-05).

ies for more general efforts to investigate sociolinguistic aspects of language change (e.g., Sinem aki and Ahola, 2023).

Despite a large number of studies on theoretical and typological aspects of AMG, (linguistically annotated) corpora are rather rare. Sampanis and Prokopidis (2021) is the first (and to our knowledge, the only other) study where a treebank of AMG is developed. However, the treebank is not yet released to the public.⁴ Some linguistic data collection efforts through fieldwork were reported in the literature (e.g., Dawkins, 1916; Bagri ık, 2018). Nevertheless, to our knowledge, no other linguistically-annotated corpora of AMG dialects exists currently.

3. Treebank and Annotation Process

The sentences in the treebank come from a collection of fairy tales in various Cappadocian (sub)dialects by Papadopoulos (2012). The collection includes fairy tales in different AMG dialects, such as Phrasiot, Silli, Aravan and Gourdounos. The tales were collected by Papadopoulos (2012) through interviews with dialect speakers in Greece. Even though our long-term objectives is building a comparable set of treebanks for all AMG dialects, for consistency, we currently annotate fairy tales only in Phrasiot dialect.⁵ Namely, we annotate the following six tales:

-  ων τζαναβαρίων το μεντζιλίσι ‘The meeting of the wild beasts’
- Αωπός τζαι το λαχτόρι ‘The fox and the rooster’
- Αωπός τζαι το λαχτόρι Β ‘The fox and the rooster Β’
- Του Βασιλό τα φσάχα τζ’ οι Ντιλπέρτσες ‘The kings sons and the Beauties’
- Το σοιρίδι τζ’ο ασλάνος ‘The pig and the lion’
- Δουλγέρ, Κουγιουμπζής, Τερζής τζαι τ’ ένα καό Θεού νομάτ ‘The carpenter, the jeweller, the tailor and a good man of God’

These tales amount to a total of 350 sentences with 4118 tokens, with an average of 11.77 tokens per sentence. The sentences were annotated by two annotators (first two authors of this paper),

⁴The treebank is listed as an ‘upcoming treebank’ on the UD repository https://github.com/UniversalDependencies/UD_Cappadocian-AMGiC/. However, no data is published as of March 2024.

⁵For an in-depth description of the Phrasiot Greek, please see Bagri ık (2018).

both native Greek speakers with linguistics degrees, and with prior exposure to Ancient Greek and Greek dialectology. Except for one of the fairy tales, all annotations (sentence segmentation, tokenization, part of speech tagging, morphological features, and dependency labels) were done manually. We also include an English translation of each sentence (translated by the annotators), and automatic transliterations to Latin script through the Python `transliterate` library.⁶ Morphosyntactic annotations were done using UD Annotatrix (Tyers et al., 2017). All sentences except 25 sentences were annotated singly, although the annotators worked closely, discussing the difficult annotation decisions. 25 sentences (313 tokens) were annotated independently by both authors for POS tags and dependency relations. On these doubly-annotated sentences the annotators agreed on 87.54% of POS tags (Cohen’s κ : 0.8581), 70.93% of dependency labels (disregarding the matching head, Cohen’s κ : 0.6841). We observed 80.83% agreement on head identification, and 60.38% agreement on full dependency labels (head and relation label). After identifying and resolving the disagreements, the annotators revised the earlier annotations, and corrected the errors due to disagreements.

The annotators mainly relied on UD general guidelines and linguistic description in Dawkins (1916) during annotation. The annotation process also benefited from documentations of earlier Standard Modern Greek treebanks (Prokopidis et al., 2005; Prokopidis and Papageorgiou, 2014, 2017). Our main annotations are compatible with the existing standard Greek treebanks in UD, facilitating the use of these treebanks for comparative research. As a result, our treebank is valid according to standard Greek validation rules with the exception of a few additional auxiliaries. These auxiliaries include the words: α ‘will’, $\chi\alpha$ ‘will’, $\theta\alpha$ ‘will’, $\pi\rho\acute{\epsilon}\phi\tau\epsilon\iota$ ‘have’ and $\acute{\epsilon}\nu\iota$ ‘am’.

We also experimented with the use of UDPipe 2 (standard) Greek model (Straka, 2018) trained on UD_Greek-GDT treebank as a pre-annotation step, after which the annotators have corrected the automatic annotations instead of annotating from scratch. Although we did not assess the differences systematically, annotators found the resulting workflow somewhat faster. However, as expected, they also indicated many errors at POS level, particularly for words that do not exist in Modern Standard Greek. The POS errors also proliferated to the automatic dependency annotations. Another observation of common errors is the gender assignment, which may be due to the loss of grammatical gender distinctions in AMG known

⁶<https://pypi.org/project/transliterate/>.

from earlier literature (Dawkins, 1916).

Following earlier Greek treebanks, we segment adposition–determiner contractions, e.g., $\sigma\tau\omicron$ ‘in the’ is tokenized as two syntactic words, σ ‘in’ and $\tau\omicron$ ‘the’. There are 51 instances of multi-word tokens. We use all UD POS tags.⁷ Similarly, we use all UD dependency relations except `clf`, `csubj`, `compound`, `dep`, `fixed`, `flat`, `goeswith`, `list`, `orphan` and `reparandum`. all the other dependency relations are attested inside the treebank. Currently, we do not use any language-specific relation labels. Some of the missing tags or relations are due to the language (e.g. `clf`), and genre (e.g., `list`) while others are likely due to the small size of the corpus (e.g., `csubj`). The morphological annotations include `Case`, `Gender`, `Number` for adjectives, determiners, nouns and pronouns, `Aspect`, `Mood`, `Number`, `Person`, `Tense`, `VerbForm` and `Voice` for verbs. The determiners and pronouns are also marked for `PronType`, and adverbs also include the `Degree` feature. A full listing of tags and relations used, as well as their frequency will be listed in the Appendix A.

4. Discussion

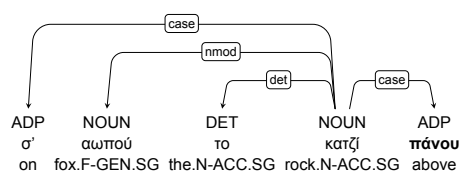


Figure 2: An example phrase from our treebank with adposition following the noun phrase ‘above the rock of the fox’.

One of the benefits of a linguistically annotated corpus is providing direct evidence for linguistic hypotheses from naturally occurring examples of text or speech, and supporting the earlier studies by quantitative means. In this section we discuss a few brief notes on some linguistic phenomena observed in our treebank by means of examples and statistical analysis.

One of the well known aspects of the AMG dialects is the tendency towards agglutinative inflections rather than more fusional morphology of standard Greek demonstrated in (1), an example from our treebank based on the schema developed by (Janse, 2019).

⁷The definitions of the tags and relation labels can be found in UD annotations guidelines at <https://universaldependencies.org/guidelines.html>.

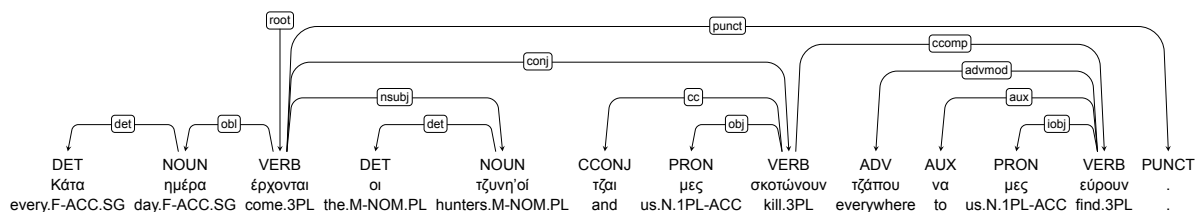


Figure 3: An example sentence ‘Every day the hunters come and kill us everywhere they find us.’ from the treebank with verbs having their objects to their left.

- (1) a. *ρνιθί -ου -v*
 hen.N -GEN -PL
 ‘of (the) hens’ [AMG (Pharasiot)]
- b. *ρνιθί -ων*
 hen.N -GEN.PL
 [Standard Modern Greek]

Instead of the fused morpheme for *-GEN.PL* in (1b), the AMG dialects tend to use clearly separate affixes for each morpheme as in (1a), which is similar to Turkish *tavuk-lar-in* ‘hen-PL-GEN’. Further similar examples are observed in our treebank.⁸

Previous studies observed that AMG dialects tend to neutralize the gender distinction (Revithiadou et al., 2017) – again, a probable contact-induced phenomenon due to Turkish which does not make any gender distinctions. We compare the ratio of common nouns with neutral gender in our treebank and the other Greek treebanks. In our treebank, 58.62% of the common nouns have neutral gender, as opposed to 28.07% in modern standard Greek treebanks, and 17.69% in ancient Greek treebanks.⁹

A common contact-induced syntactic property of Pharasiot, as well as other AMG dialects, is the placement of adpositions. Rather than prepositions as in standard Greek, AMG dialects place some of the adpositions after the noun phrase they modify. Figure 2 demonstrates this with an example from the treebank. The same phrase in standard Greek (*Πάνω στον βράχο της αλεπούς* ‘On the rock of the fox’) places the adposition before the noun phrase. Again, in Pharasiot the placement of adposition shows similarities with Turkish (*tilkinin taşının üstünde*), which almost exclusively makes use of postpositions.

Looking at the rate of postposed adpositions, 16 of 240 of the adpositions follow the NP, with a rate of 6.25%. This shows that the sample of Pharasiot

⁸Precise quantification would require morpheme segmentation, which is beyond the scope of the current study.

⁹All statistics were reported are based on UD release 2.12 of GDT and GUD for modern Greek, and Perseus (Bamman and Crane, 2011) and PROIEL (Haug and Jøhndal, 2008) for ancient Greek.

in our treebank predominantly makes use of prepositions. However, looking at the same ratio on both existing treebanks, the ratio of postpositions is only 0.54%, for modern standard Greek treebanks and 3.08% the Ancient Greek treebanks. Although these numbers should be interpreted with a grain of salt due to small treebank size and genre differences, they confirm that the AMG dialect has the tendency to postpose the adpositions. Closer inspection of these differences may provide further evidence on the source of differences between different modern Greek dialects.

Finally, we report some statistics regarding basic word order. Looking at the verbal predicates of the matrix clause, 23.62% of the *obj* or *ccomp* relations follow the verb, diverging from the dominant SVO pattern. This is only slightly larger than the modern standard Greek with main verbs having 19.50% their objects to their left. Here, ancient Greek treebanks provide a more divergent left-object ratio with 45.87%. The ratio of subjects (*nsubj* or *csbj*) to the right of their verbs, on the other hand, is 52.03%, in comparison to 26.43% for modern Greek, and 29.94%, ancient Greek treebanks. This is a much stronger divergence from the SVO order, indicating a more flexible word order. However, as before, it needs to be interpreted with care since the genre (fairy tales) are likely to be an important factor.

5. Conclusions and Outlook

This paper presented a treebank for a critically endangered language variety, a modern Greek dialect of Asia Minor, with interesting linguistic properties due to historical development and linguistic contact. Due to the small number of native speakers to consult with, relatively scarce language documentation, and shortage of the amount and the variety of resources available, building such treebanks requires a great deal of effort. However, the effort is worthwhile as the treebank complements linguistic descriptions like grammars and data from fieldwork, and provides more objective, quantitative means of studying the language.

We also presented example analyses investigating some of the linguistic phenomena discussed in

the recent literature, using examples and statistics over the treebank. Although care is needed for interpreting the results due to the modest size and restricted genre, they nevertheless showcase the benefits of treebanks for studying low resource languages.

The AMG treebank project has ample room for growth and expansion. One of the future directions we consider is expanding the treebank by adding more sentences and texts, including diverse genres beyond fairy tales and other (sub)dialects. This will provide a broader representation of the AMG dialects and hopefully incite some interest from the community (AMG dialects are studied by a relatively small but active group of researchers) for collaborating and creating a larger corpus. With their collaboration we could, for example, add annotations for contact-induced linguistic phenomena (e.g., using the taxonomy introduced by [Sampanis and Prokopidis, 2021](#)) and create treebanks from parallel or comparable source material in multiple dialects, or material collected before and after forced migration of AMG speakers.

6. Limitations

The main limitation of the present work is the relatively small size of the treebank, with small diversity in the text types. However, we believe this is a limitation of the nature of building treebanks for such low resource languages.

7. Ethical Considerations

Our source material comes from fairy tales. As a result, we do not believe they contain any privacy problems.

Although it is unlikely for fairy tales, folk tales, reflecting the ‘wisdom’ of the times they have survived through, may contain some discriminating expressions of biases. As a result care may be needed for tools or resources developed using this treebank.

8. Acknowledgements

We thank the three anonymous reviewers for their insightful comments and suggestions.

9. Bibliographical References

- Nikolaos P. Andriotis. 1948. *Το γλωσσικό ιδίωμα των Φαρασών [The dialect of Pharasa]*. Ikaros, Athens.
- Metin Bagriaçık. 2018. *Pharasiot Greek: word order and clause structure*. Ph.D. thesis, Ghent University.
- David Bamman and Gregory Crane. 2011. The ancient Greek and Latin dependency treebanks. In *Language technology for cultural heritage: Selected papers from the LaTeCH Workshop Series*, pages 79–98. Springer.
- Richard M. Dawkins. 1916. *Modern Greek in Asia Minor, a study of the dialects of Silli, Cappadocia and Phârasa*. Cambridge University Press, Cambridge.
- Marie-Catherine De Marneffe, Christopher D Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational linguistics*, 47(2):255–308.
- Harald Hammarström, Robert Forkel, Martin Haspelmath, and Sebastian Bank, editors. 2024. *Glottolog 5.0*. Max Planck Institute for the Science of Human History, Jena.
- Dag TT Haug and Marius Jøhndal. 2008. Creating a parallel treebank of the old Indo-European Bible translations. In *Proceedings of the second workshop on language technology for cultural heritage data (LaTeCH 2008)*, pages 27–34.
- Mark Janse. 2004. Animacy, definiteness, and case in Cappadocian and other Asia Minor Greek dialects. *Journal of Greek linguistics*, 5(1):3–26.
- Mark Janse. 2009. The resurrection of Cappadocian (Asia Minor Greek). *Αω International*, 11:10–15.
- Mark Janse. 2019. Agglutinative noun inflection in Cappadocian. In *The Morphology of Asia Minor Greek*, pages 66–115. Brill.
- Mark Janse. 2020. Back to the future: Akritic light on diachronic variation in Cappadocian (east Asia Minor Greek). *Variation in Post-classical and Byzantine Greek*, pages 201–239.
- Petros Karatsareas. 2011. *A study of Cappadocian Greek nominal morphology from a diachronic and dialectological perspective*. Ph.D. thesis, University of Cambridge.
- Petros Karatsareas. 2016. Convergence in word structure: Revisiting agglutinative noun inflection in Cappadocian Greek. *Diachronica. International Journal for Historical Linguistics. Founded by EFK Koerner, General Editor, 1984–2001*, 33(1):31–66.
- Nikolaos G. Kontossopoulos. 1981. *Διαλεκτοί και ιδιώματα της Νεας Ελληνικής [Dialects and idioms of Modern Greek]*. Grigoris, Athens.

- Dimitra Melissaropoulou and Angela Ralli. 2020. Revisiting the borrowability scale (s) of free grammatical elements: Evidence from Modern Greek contact induced varieties. *Journal of language contact*, 12(3):707–736.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. [Universal Dependencies v1: A multilingual treebank collection](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666, Portorož, Slovenia. European Language Resources Association (ELRA).
- Iordanis B. Papadopoulos. 2012. *Καππαδοκικά Παραμυθία & Λαϊκές Παραδόσεις: Φαρασα, Σιλλή, Αραβαν, Γουρδουνος [Cappadocian stories and lore: Pharasa, Silli, Aravan and Ghourzono]*. KE.MI.PO, 2nd Edition.
- Prokopis Prokopidis, Elina Desipri, Maria Koutsombogera, Harris Papageorgiou, and Stelios Piperidis. 2005. Theoretical and practical issues in the construction of a greek dependency corpus. In *Proceedings of the 4th Workshop on Treebanks and Linguistic Theories (TLT-2005)*.
- Prokopis Prokopidis and Haris Papageorgiou. 2014. [Experiments for dependency parsing of Greek](#). In *Proceedings of the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages*, pages 90–96, Dublin, Ireland. Dublin City University.
- Prokopis Prokopidis and Haris Papageorgiou. 2017. [Universal Dependencies for Greek](#). In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 102–106, Gothenburg, Sweden. Association for Computational Linguistics.
- Angela Ralli. 2009. Morphology meets dialectology: insights from Modern Greek dialects. *Morphology*, 19:87–105.
- Angela Ralli. 2020. Matter versus pattern borrowing in compounding: Evidence from the Asia Minor Greek dialectal variety. *Morphology*, 30(4):423–446.
- Anthi Revithiadou, Marc Oostendorp, Kalomoira Nikolou, and Maria-Anna Tiliopoulou. 2006. Vowel harmony in contact-induced systems: the case of Asia Minor dialects of Greek. In *Proceedings of the 2nd international conference of Modern Greek dialects and linguistic theory, September 30–October 3, 2004*, pages 350–365.
- Anthi Revithiadou, Vassilios Spyropoulos, and Giorgos Markopoulos. 2017. From fusion to agglutination: The case of Asia Minor Greek. *Transactions of the Philological Society*, 115(3):297–335.
- Konstantinos Sampanis and Prokopis Prokopidis. 2021. [Asia Minor Greek in contact \(AMGiC\): Towards a dialectal treebank comprising contact-induced grammatical changes](#). In *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)*, pages 86–95, Sofia, Bulgaria. Association for Computational Linguistics.
- Kaius Sinnemäki and Noora Ahola. 2023. [Testing inferences about language contact on morphosyntax: A typological case study on alorese–adang contact](#). *Transactions of the Philological Society*, 121(3):513–545.
- Milan Straka. 2018. [UDPipe 2.0 prototype at CoNLL 2018 UD shared task](#). In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 197–207, Brussels, Belgium. Association for Computational Linguistics.
- Francis M. Tyers, Mariya Sheyanova, and Jonathan North Washington. 2017. [UD Annotatrix: An annotation tool for Universal Dependencies](#). In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, pages 10–17, Prague, Czech Republic.

A. Further treebank statistics

Feature	Frequency
Number	2362
Case	1774
Gender	1733
PronType	930
Person	684
Aspect	591
VerbForm	591
Voice	591
Mood	589
Tense	553
Definite	401
Poss	57
NumType	15
Foreign	13

Table 1: Morphological features and their frequencies.

POS Tag	Frequency
PUNCT	728
NOUN	683
DET	658
VERB	591
PRON	326
ADP	256
CCONJ	172
ADV	162
AUX	158
SCONJ	152
ADJ	60
NUM	58
PART	55
INTJ	39
X	12
PROPN	7
SYM	1

Table 2: POS tag distribution.

Dependency label	Frequency
punct	728
det	618
root	350
obl	238
case	234
obj	229
nsubj	219
conj	195
nmod	189
advmod	188
mark	174
cc	171
aux	113
amod	57
parataxis	55
nummod	46
cop	45
vocative	42
advcl	37
iobj	37
discourse	32
ccomp	31
xcomp	29
expl	16
dislocated	13
appos	8
nsubj:pass	8
acl:relcl	5
acl	4

Table 3: Dependency label distribution.