Medieval Social Media: Manual and Automatic Annotation of Byzantine Greek Marginal Writing

Colin Swaelens[1], **Ilse De Vos**[2] and **Els Lefever**[1]
[1] LT3, Language and Translation Technology Team
Department of Translation, Interpreting and Communication
[2] Department of Linguistics Ghent University, 9000 Ghent, Belgium
{colin.swaelens, i.devos, els.lefever}@ugent.be

Abstract

In this paper, we present the interim results of a transformer-based annotation pipeline for Ancient and Medieval Greek. As the texts in the Database of Byzantine Book Epigrams have not been normalised, they pose more challenges for manual and automatic annotation than Ancient Greek, normalised texts do. As a result, the existing annotation tools perform poorly. We compiled three data sets for the development of an automatic annotation tool and carried out an inter-annotator agreement study, with a promising agreement score. The experimental results show that our part-of-speech tagger yields accuracy scores that are almost 50 percentage points higher than the widely used rule-based system Morpheus. In addition, error analysis revealed problems related to phenomena also occurring in current social media language.

1 Introduction

Despite the nonexistence of the world wide web in the Middle Ages, Byzantine book epigrams bear some resemblance to current social media, such as Twitter.[1] Just like a tweet, a book epigram is usually a rather short, personal statement of an author, who expresses themselves on their daily occupation, i.e. copying manuscripts. Furthermore, the typeface of both tweets and book epigrams displays a lot of *orthographic inconsistencies* as the content is often written phonetically. However, the big difference between social media and Byzantine book epigrams is the amount of text available for NLP: 575,000 tweets are sent every minute, while the Database of Byzantine book epigrams (DBBE) (Ricceri et al., 2023) counts 12,000 epigrams in total.

The Byzantine book epigrams that make up the DBBE, can be defined as metrical paratexts, i.e. poems standing next to (para, from the Greek word παρά) another text or figure. They often appear in the margins of manuscripts or as scribblings between two paragraphs. Concerning content, these epigrams, among other things, comment on the main text of the manuscript, give some insight in the life of the scribe or show off the scribe's knowledge. DBBE Occurrence 32143 serves as an example, provided with the authors' translation:

(1) ὤσπερ ξένοι χαίρουσιν ἰδεῖν πατρίδα
οὕτω καὶ οἱ γράφοντες βιβλίου τέλος
Just like travellers rejoice upon seeing their homeland,
so do writers upon reaching the end of a book.[2]

The orthographic idiosyncrasies these book epigrams display are mainly due to a phonetic evolution, called *itacism*, which indicates the shift of the classical Athenian pronunciation of the vowels ι [i], η [ɛ], υ [y] and the diphthongs ει [ɛj], οι [oj] to the pronunciation [i]. The scribe of the book epigram – who may or may not have authored it – did not always know (or care?) which of the five [i]'s needed to be written. The disyllabic word ἰδεῖν (to look), for example, is present in 19 different forms in DBBE. Exactly that is the added value of DBBE compared to other pre-Modern Greek corpora: these corpora generally provide Greek that is normalised to an Ancient Greek model, while DBBE provides both the original transcription of the manuscript and an edited, normalised version. The former is called *occurrence*, the latter *type*.

[1] *Byzantine* and *Medieval* will be used as synonyms, covering the period between ca. 500 and 1500 AD

[2] Translations are made by the authors, unless stated otherwise.

Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII), pages 1–9
July 13, 2023 ©2023 Association for Computational Linguistics

The texts of the DBBE will be subject of further linguistic and literary research, for which these texts are ideally all annotated. Since manual annotation is not feasible for all words, we opted for an automatic way to do so. Preliminary tests showed that existing systems for morphological analysis do not perform well on the text of the occurrences. To overcome the shortcomings of current systems for morphological analysis, we developed a novel transformer-based part-of-speech tagger for Ancient and Medieval Greek.[3] To evaluate the performance of the tagger, a novel gold standard for Byzantine Greek was developed, where all tokens were provided with a coarse-grained part-of-speech tag and full morphological analysis. In addition, we also performed an error analysis, which revealed several problems that are very typical to this kind of texts, i.e. texts where the material context (the manuscript) strongly affects the language.

2 Related Research

The interest in NLP for pre-Modern Greek has increased over the last few years, thanks to – among other things – the availability of open-source corpora. The first corpus initiative for Greek texts was the Thesaurus Linguae Graecae (TLG) (Pantelia, 2022), a comprehensive digital library of Greek texts written between 800 BC and 1453 AD (viz. the fall of Byzantium), that sums up to more than 110M tokens, covering 10,000 works and 4,000 authors. The TLG, however, is not freely available. An open-source alternative is the Open Greek and Latin Project[4], that consists of the Perseus Digital Library (Crane, 2022), a collection of more than 13,5M tokens of mostly classical Greek prose and poetry, on the one hand, and the First1K Project, a complementary part to Perseus summing up to 25,5M tokens of classical and post-classical Greek prose and poetry[5].

In addition to these two text corpora, several treebanks were developed. The Ancient Greek Dependency Treebank (AGDT) (Bamman and Crane, 2011; Celano, 2019) stores 560,000 tokens from both classical prose and poetry, that were manually provided with a part-of-speech tag, morphological analysis, lemma and syntactic relation. PROIEL (Haug and Jøhndal, 2008) has a more specific content: the treebank stores the New Testament in Greek and four other languages, counting 277,000 tokens. The Gorman treebank (Gorman, 2020) is a treebank of around 550,000 tokens of exclusively classical Greek prose. As a last example, the Pedalion Trees (Keersmaekers et al., 2019) are almost completely complementary to the AGDT (apart from some texts) and count some 320,000 tokens. The Pedalion Trees contain annotated texts from Trismegistos (Depauw and Gheldof, 2014), a database of papyrus texts, that displays the original text with all its idiosyncrasies and even *errors*, just like the occurrences in DBBE. All of the above mentioned treebanks make use of or have extended the Universal Dependencies (Nivre et al., 2017).

Since the development of Morpheus (Crane, 1991), a rule- and dictionary-based system to perform part-of-speech tagging (or morphological analysis) of Greek tokens, multiple part-of-speech taggers have been developed to cope with Morpheus' two main pitfalls: it does not disambiguate ambiguous forms and it cannot deal with out-of-vocabulary words. Celano et al. (2016) did a comparative study, which showed that MateTagger (Bohnet and Nivre, 2012) outperformed Hunpos tagger (Halácsy et al., 2007), RFTagger (Schmid and Laws, 2008), the OpenNLP part-of-speech tagger[6] and NLTK Unigram tagger (Bird, 2006) on Ancient, normalised Greek data. When Keersmaekers (2019) repeated that experiment with Mate tagger, RFTagger and MarMot tagger (Mueller et al., 2013) to find out which is best suited for papyrological data, RFTagger outperformed the other two. Schmid (2019) also developed RNN tagger, the neural counterpart of RFTagger. Singh et al. (2021) explored the possibilities of a transformer-based part-of-speech tagger on DBBE types, the normalised text of the book epigrams, which yielded promising results.

[3] As Greek is a highly inflectional language, we use part-of-speech tag to cover both the part-of-speech and the full morphological analysis of a word in the rest of the paper.

[4] https://opengreekandlatin.org

[5] https://opengreekandlatin.github.io/First1KGreek/

[6] https://opennlp.apache.org

3 Data Compilation and Annotation

Our aim is not to annotate the DBBE *types*, the normalised poems, but the DBBE *occurrences*. To achieve this, we trained a transformer-based language model, of which the embeddings are used to train a part-of-speech tagger. Section 3.1 describes the data sets used for training the language model and fine-tuning it for part-of-speech tagging, while Section 3.2 describes the manual annotation and validation of the Byzantine Greek evaluation set.

3.1 Training Data Compilation

Since transformer-based language models are very greedy and the Greek data available is rather scarce, we complemented all corpora described in Section 2, except for the TLG, with the Modern Greek Wikipedia data, shown in Figure 1. This is done, because Byzantine Greek is situated in time between Ancient Greek and Modern Greek, and because Byzantine Greek displays already quite some Modern Greek characteristics (Holton et al., 2019). Data labelled as *incerta* could not be situated in any time period, *varia* treats anthologies. From now on, we call this the LM data set. In addition to this data set, we compiled a training set for the part-of-speech tagger, consisting of all above described treebanks, summing up to 1,132,120 Ancient Greek tokens.

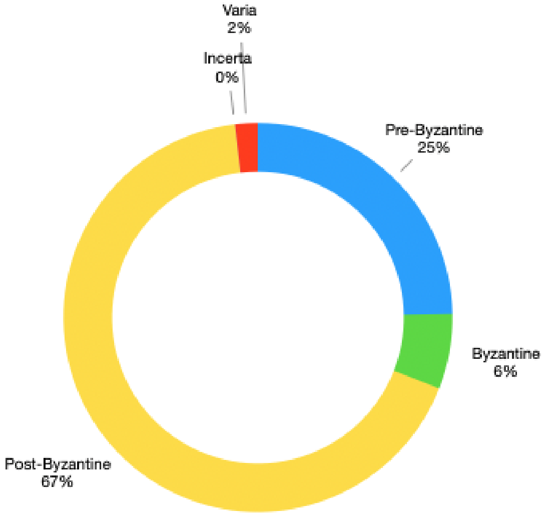

Figure 1: BERT training data

3.2 Evaluation Data Annotation

We compiled a test set of 10,000 tokens from the DBBE *occurrences* to evaluate whether the part-of-speech model is able to analyse the Byzantine data given its training on Ancient Greek data. This evaluation set has been manually annotated, following the AGDT annotation guidelines (Celano, 2018), so that the DBBE, when eventually annotated, is complementary to the existing resources. However, we first carried out an inter-annotator agreement experiment (IAA), which has – to the best of our knowledge – not yet been conducted for either Ancient or Byzantine Greek. The aim of this IAA study is twofold: firstly to evaluate whether the label set shown in Table 1 is suitable for this corpus of Byzantine book epigrams; secondly to evaluate whether the manual annotations are reliable and consistent across annotators, which is a prerequisite to use the resulting corpus for evaluating and – in the near future – training our part-of-speech tagger.

Given the nature of our data, we saw it necessary to add one label to the AGDT label set, namely *missing*. As mentioned above, our corpus consists of faithful manuscript transcriptions. As shown in Example 2, words or word groups that are illegible are marked with (...). These so-called lacunae are rather rare in Greek text editions. This is why pre-existing corpora – which consist only of text editions – do not need any label for them. For easy reference, we decided to name this label *missing*.

(2) *(...) χρόνον τε και λόγους καὶ τὴν*
(...) χronon te kje logus kje tin
φύσιν
fisin

(...) time and also words and the nature

DBBE Occurrence 30520

The IAA experiment was carried out by three annotators, linguists with profound knowledge of Ancient Greek. They were asked to annotate some 1,000 tokens we extracted from the epigrams shown in Table 3 with the features shown in Table 1. Because of the highly inflectional nature of the Greek language, the annotation consisted of both the assignment of a part-of-speech and the token's morphological analysis. Since the part-of-speech tag and the morphological analysis of a token are aggregated in one label, our tag set sums up to more than 1,200 labels. The eventual tag consists of nine slots, corresponding to the nine columns in Table 1. This label set follows, just like the treebanks in Section 2, the Universal Dependencies label set. To relieve the annotators, we bootstrapped the tokens making use of Singh et al.'s part-of-speech tagger to already suggest a morphological analysis. However, this was a difficult assessment, as we know that the annotators might be influenced by the result of the bootstrapping. The annotators were asked to annotate no more than two hours a day to assure that they could stay focused. Upon completion, we calculated the IAA scores with Fleiss' Kappa.

The IAA experiment resulted in an agreement of 92.72% for the part-of-speech and 89.83% for the complete morphological analysis. The agreement scores are very high, showing *almost perfect* agreement (>90%) for the part-of-speech tagging and morphological analysis in isolation, and very *strong* agreement (80-90%) for the combined label. These scores are very encouraging, especially because we perform part-of-speech tagging on Greek data, for which different tags are often possible and arguments can be made for different analyses of the same word.

This can be illustrated with the word χάριν (*on behalf of*) followed by a genitive. One can argue that its part-of-speech is a noun, χάρις, since its accusative is used in an adverbial way. It is just as valid, however, to state that χάριν is an adverb *an sich*. In our test set, not once is there an agreement between the three annotators about the token χάριν. One of the annotators consistently tags χάριν as a preposition, while the other two annotators tagged two occurrences of χάριν as noun, and the other four as preposition. For the eventual annotation, χάριν is tagged as adverb when followed by a genitive; otherwise it is tagged as a noun.

While further investigating cases of disagreement, some tendencies caught the eye. About 50% of the disagreement is attributed to the part-of-speech tag, especially the difference between *noun* and *adjective*. According to the dictionary LSJ (Liddell et al., 1966), the last word of Example 3, φίλον (friend), is an adjective. This adjective, however, can be substantivised by putting an article in front, as is the case in Example 3. Two of our annotators tagged φίλον as an adjective, one as a noun. For the eventual annotation of the gold standard, these substantivised adjectives were annotated as a noun.

(3) χείρας ἐκτείνας δεξιοῦται τον φίλον
with extended hands, he greets his friend
DBBE Occurrence 21375

The next category of disagreement is related to the gender of words. Quite some Greek words have the same morphology for both masculine and feminine, e.g. the adjective ἄπιστος (*untrue*), or for both masculine and neutral, e.g. the genitive singular ἀγαθοῦ (*good*), or even for the three genders, e.g. the article in the genitive plural τῶν (the). The article τῶν is twelve times attested in our IAA study and caused disagreement four times. In our view, this is due to fatigue or negligence of the annotators, as the gender can be deducted from the agreeing noun, as shown in Example 4. Two annotators tagged this τῶν as masculine, notwithstanding its agreement with the neutral word βουλευμάτων (*decisions*).

(4) ἐπήβολος φρὴν τῶν σοφῶν βουλευμάτων
the intelligence, partaking in wise decisions
DBBE Occurrence 30520

For future annotations we explicitly pointed out to not assign a tag before the whole constituent was read, in the hope to prevent this type of inaccuracies.

Nevertheless, we dare say that the label set is well suited for this annotation task, given the high agreement scores.

4 A Novel Part-of-Speech Tagger for Byzantine Greek

4.1 BERT Language Model

As we desire our part-of-speech tagger copes with all idiosyncrasies of our Medieval Greek corpus, the need emerged to include context

PoS	Person	Number	Tense	Mood	Voice	Gender	Case	Degree
adjective	1	singular	aorist	imperative	active	common	nom	comp
adverb	2	plural	future	indicative	medial	feminine	acc	super
article	3	dual	fut. perf.	infinitive	med-pass	masculine	gen	-
conjunction	-		imperfect	optative	passive	neutral	dat	
exclamation			perfect	participle	-	-	voc	
interjection			pluperfect	subjunctive			-	
punctuation			present	-				
noun			-					
numeral								
particle								
preposition								
pronoun								
verb								
missing								

Table 1: Overview of the nine slots that make up the part-of-speech tag of each token. That tag is a combination of the part-of-speech and the morphological analysis of the token.

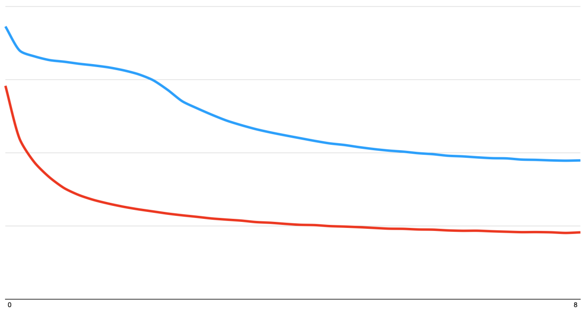

Figure 2: Convergence of loss on held out test set. The blue graph is the pre-Byzantine and Byzantine data set, the red one is complemented with post-Byzantine greek.

into the model. Firstly, we developed two BERT (Devlin et al., 2018) language models: one that has been trained on the LM data set *without* Modern Greek, described in Section 3.1, and a second that has been trained on the complete LM data set, including Modern Greek. This LM data set consists of 31,467,014 pre-Byzantine tokens, 7,952,719 Byzantine tokens, 85,575,140 post-Byzantine tokens and 2,418,672 tokens that could not be classified in one of the previous classes, counting 127,413,536 tokens in total, as shown in Figure 1. This data served as input for the BERT model, optimised for Masked Language Modelling, with the following parameters: 15% of the input tokens are replaced by [MASK] tokens, the maximum sequence length per sentence was limited to 512 sub-words and 12 hidden layers were used. The validation loss convergence as a function of time of both language models is shown in Figure 2.

As illustrated by the loss functions in Figure 2, it is clear that the language model trained on all pre-Byzantine, Byzantine and post-Byzantine Greek data performs best. We call this language model DBBErt, and made it available for the research community[7]. This model will be the basis for the fine-tuning for part-of-speech tagging.

4.2 Part-of-Speech Fine-tuning

As a second step, the DBBErt language model is incorporated into our part-of-speech tagger, that, as mentioned in Section 1, also provides the full morphological analysis.

As a training set, we used the treebanks described in Section 2 and extracted the part-of-speech tags and morphological information. In addition we extended the training set with 2,000 manually annotated tokens from DBBE occurrences. To train the part-of-speech tagger, we made use of the FLAIR framework (Akbik et al., 2019). The contextual token embeddings from DBBErt (cf. Section 4.1) are stacked with randomly initialised character embeddings. These are processed by a bidirectional long short-term memory (LSTM) encoder and a conditional random field (CRF) decoder: a combination commonly used for sequential tagging tasks. The LSTM has a hidden size of 256 and starts with a learning rate of 0.1 that is linearly decreased during training.

4.3 Evaluation of the Part-of-Speech Tagger

For the evaluation of our part-of-speech tagger, we have to keep in mind that the training was

[7] This model is available at https://huggingface.co/colinswaelens/DBBErt

Model	Accuracy	Precision	Recall	F1
RNN Tagger	63.04%	65.27%	63.04%	61.92%
fine-tuned pre-Byzantine and Byzantine LM	63.29%	69.19%	63.29%	62.14%
fine-tuned DBBErt LM	69.89%	73.22%	68.57%	67.32%

Table 2: Evaluation scores for the full morphological analysis for (1) RNN Tagger, (2) the tagger fine-tuned on the LM containing pre-Byzantine and Byzantine data, and (3) the tagger fine-tuned on DBBErt (containing all Greek data).

done with mostly Ancient, normalised Greek data, while the evaluation set existed of not-normalised Byzantine Greek epigrams. As an intermediate step, we first evaluated the performance of our part-of-speech tagger on a test set consisting of manually annotated tokens from DBBE types. Our model yielded an accuracy score of 83.64%, a score competitive to Singh et al.'s 86.66% on that same test set. The slight difference in performance might be attributed to the fact that Singh et al. retrained a Modern Greek language model that stripped off all diacritics of both training and test data. Our model, however, did take into account all diacritics present in Medieval Greek.

The final evaluation, however, is performed on 8,000 tokens from DBBE occurrences and resulted in 69,89% accuracy. The drop in accuracy is not surprising, given the very challenging nature of the Byzantine poems, which is also illustrated by the performance of Morpheus (cf. Section 2) on our test set of occurrences. Morpheus could not process 44% of the test set (out-of-vocabulary tokens), 30% of the tokens were ambiguous and not disambiguated, while only 24% of the test set was disambiguated. In the end did Morpheus yield an accuracy score of 19%. We also compared our results with RNN Tagger (Schmid, 2019), a neural model that displayed state-of-the-art results for Ancient Greek. As shown in Table 2, our novel part-of-speech model outperforms RNN tagger, which obtains an accuracy score of 63%, with more than 6 percentage points. For completion, we also trained a model fed with the word embeddings from the smaller pre-Byzantine and Byzantine model. This model, which was not trained on post-Byzantine data, clearly performs worse that the tagger fine-tuned on the full language model. In addition to this quantitative analysis, we also performed a qualitative analysis of the results of our part-of-speech tagger with a special focus on two phenomena that also appear in current social media posts.

5 Error Analysis and Discussion

As mentioned in Section 1, the book epigrams bear some resemblance to modern social media posts. Exactly those similarities are an interesting starting point for our error analysis.

Let us begin with the appearance of a social media comment, which can accompany a picture, an opinion on someone's message, or just a retweet of another tweet. Those social media comments could be categorised as paratexts: a text standing next to (para, from the Greek παρά) another text, just as our book epigrams. They are mostly to be found in the margins around the main text of a manuscript, a material property of this corpus that determines the first category of errors. To illustrate this error type, we will discuss the following verse (English translation in italics):

(5) + X(ριστὸ)ν ἀεὶ ζώοντα θεβροτὸν αὒτὸν ὄντα :·
Christ, the always living God, being mortal as well.
DBBE Occurrence 20483

The word *θεβροτὸν* is not an existing word but a mistake made by the scribe, who erroneously combined the abbreviation of θεόν (God) with the next word, βροτόν (mortal). Although it was standard practice to abbreviate θεόν as θε with a dash above it, it is clear that the scribe of this manuscript did not intend to write an abbreviation. 146 related occurrences show that our scribe did not realise this was an abbreviation, and thus wrote the two words as a compound. The performance of the part-of-speech tagger, however, was not affected too much by these irregularities: θεβροτὸν was analysed as a noun, accusative masculine singular, which is the correct analysis of βροτὸν. Most of the other erroneous compounds are analysed correctly, what might be attributed to the sub-word tokenizer used to train DBBErt. The opposite phenomenon, erroneously split words, occurs as well in DBBE:

(6) νυκτα δι᾽ ἀμβροσίην τὴν οὐ θέμις ἔξον ὀμῆναι ·
Through the immortal night that should not rightfully be called by its name
DBBE Occurrence 31488

The last two words of this verse are the result of an incorrect split of the word ἐξονομῆναι. This error might have been caused by confusion with the future participle of “to have”, the existing word ἔξον, the second part, however, does not make any sense at all. Although not correctly analysed, the part-of-speech tagger made a reasonable attempt. It tagged ἔξον and ὀμῆναι as a verb, the former as active indicative aorist 3 singular, the latter as active infinitive aorist. Both analyses are, to our opinion, based on the suffixes of the words. Most of these split words are analysed incorrectly.

The second category of mistakes can be attributed to an even more salient characteristic of present social media posts, namely the writing mistakes due to a phonetic way of writing. The English word *because*, for instance, can be found on twitter as *becuz*, as both are pronounced identically. The same principle applies to a lot of words in DBBE, which are written incorrectly, as shown in the following examples:

(7) εἰρμώσας ἐζόφωσεν ἤρεν μετείχους
Being in tune, he threw it into darkness, he made an end to it with his sound [8]
DBBE Occurrence 17374

(8) ὤπο(ς) μοναστὴς νεόφυτο(ς) οἰκέτ(ης)
Thy servant the monk Neophytos[9]
DBBE Occurrence 17594

The examples above contain several spelling mistakes that were made because of a phonetic way of writing. The words εἰρμώσας and μετείχους of Example 7 are incorrect because of the itacism (See Section 1). Although the stem is completely incorrect, εἰρμώσας was analysed correctly as a verb, active participle aorist nominative masculine singular. As for μετείχους, there might be two reasons for it not being analysed correctly: the spelling mistake and the fact that it is an incorrect contraction of μετ`ἤχους. The first word of Example 8 should have been ὅπως instead of ὤπος, yet both the spiritus and the length of the vowels have lost their distinctive value after the classical period. We noticed that if the orthographic mistake happens at the ultimate and/or penultimate syllable, the algorithm outputs an incorrect morphological analysis. This is in line with our conclusion about the compound words (cf. supra): the embeddings are sub-word based, so if the sub-words are nonsensical, the part-of-speech tagger will not provide a correct morphological analysis.

6 Conclusion and Future research

The Database of Byzantine Book Epigrams stores a very challenging corpus with its own peculiarities and problems for automatic processing. This automatic processing is necessary since manual annotation is not feasible for the complete DBBE corpus. To develop a more flexible approach that is able to cope with lots of orthographic variety and out-of-vocabulary words, we trained a novel language model for Greek, the DBBErt, and fine-tuned it for part-of-speech tagging. To evaluate this part-of-speech-tagger on Byzantine Greek, we developed a novel gold standard, which was manually annotated using the AGDT annotation guidelines. This label set was first subject of an IAA study, that showed very high agreement scores.

Although the evaluation showed promising results, the error analysis exposed once more the inherent problems of the book epigrams, which philologists still agonise over.

An important next step in our research is the development of a lemmatizer, which will make the annotation of our corpus complete. Once this annotation is done, we will research how similarity can be measured between hemistichs, verses and epigrams in the DBBE, in order

[8] translation by Bentein et al. (2010)

[9] translation by Marava-Chatzēnikolaou et al. (1978)

to link similar texts copied (and sometimes altered) by different scribes.

Limitations

The main limitation of our research, is the limited amount of data available. Transformer-based language models are very data-greedy, which made us add Modern Greek data to our model for Ancient and Medieval Greek to have a substantial amount of data. The nature of the data is a second limitation. We want to process the Greek texts as they are found in manuscripts, in their original form. That entails that the texts not only contain orthographic irregularities but, as mentioned in Section 5, also words that are either erroneously split or glued together. As a result, the non-existing words in the corpus considerably impact the system performance for the task of morphological analysis.

References

Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. 2019. FLAIR: An easy-to-use framework for state-of-the-art NLP. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 54–59, Minneapolis, Minnesota. Association for Computational Linguistics.

David Bamman and Gregory Crane. 2011. The ancient greek and latin dependency treebanks. In *Language Technology for Cultural Heritage*, pages 79–98, Berlin, Heidelberg. Springer Berlin Heidelberg.

Klaas Bentein, Floris Bernard, Kristoffel Demoen, and Marc de Groote. 2010. New testament book epigrams. some new evidence from the eleventh century. *Byzantinische Zeitschrift*, 103(1):13–23.

Steven Bird. 2006. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, pages 69–72.

Bernd Bohnet and Joakim Nivre. 2012. A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1455–1465, Jeju Island, Korea. Association for Computational Linguistics.

Giuseppe G. A. Celano. 2018. Guidelines for the ancient greek dependency treebank 2.0. Last consulted December 2022.

Giuseppe G. A. Celano, Gregory Crane, and Saeed Majidi. 2016. Part of speech tagging for ancient greek. *Open Linguistics*, 2(1).

Giuseppe GA Celano. 2019. The dependency treebanks for ancient greek and latin. *Digital Classical Philology*, page 279.

Gregory R. Crane. 1991. Generating and Parsing Classical Greek. *Literary and Linguistic Computing*, 6(4):243–245.

Gregory R. Crane. 2022. Perseus digital library. Last accessed 10 February 2023.

Mark Depauw and Tom Gheldof. 2014. Trismegistos: An interdisciplinary platform for ancient world texts and related information. In *Theory and Practice of Digital Libraries – TPDL 2013 Selected Workshops*, pages 40–52, Cham. Springer International Publishing.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Vanessa B Gorman. 2020. Dependency treebanks of ancient greek prose. *Journal of Open Humanities Data*, 6(1).

Péter Halácsy, Andras Kornai, and Csaba Oravecz. 2007. Hunpos: an open source trigram tagger. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 209–212.

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.

David Holton, Geoffrey Horrocks, Marjolijne Janssen, Tina Lendari, Io Manolessou, and Notis Toufexis. 2019. *Phonology*, page 1–238. Cambridge University Press.

Alek Keersmaekers. 2019. Creating a richly annotated corpus of papyrological Greek: The possibilities of natural language processing approaches to a highly inflected historical language. *Digital Scholarship in the Humanities*, 35(1):67–82.

Alek Keersmaekers, Wouter Mercelis, Colin Swaelens, and Toon Van Hal. 2019. Creating, enriching and valorizing treebanks of Ancient Greek. In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, pages 109–117, Paris, France. Association for Computational Linguistics.

Henry George Liddell, Robert Scott, Henry Stuart Jones, and Roderick McKenzie. 1966. *A Greek-English Lexicon.* Clarendon press.

A. Marava-Chatzēnikolaou, Ethnikē Vivliothēkē tēs Hellados, C. Touphexē-Paschou, and Akadēmia Athēnōn. 1978. *Catalogue of the Illuminated Byzantine Manuscripts of the National Library of Greece: Manuscripts of New Testament texts 10th-12th century.* Catalogue of the Illuminated Byzantine Manuscripts of the National Library of Greece. Publications Bureau of the Academy of Athens.

Thomas Mueller, Helmut Schmid, and Hinrich Schütze. 2013. Efficient higher-order CRFs for morphological tagging. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 322–332, Seattle, Washington, USA. Association for Computational Linguistics.

Joakim Nivre, Daniel Zeman, Filip Ginter, and Francis Tyers. 2017. Universal Dependencies. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, Valencia, Spain. Association for Computational Linguistics.

M.C. Pantelia. 2022. *Thesaurus Linguae Graecae: A Bibliographic Guide to the Canon of Greek Authors and Works.* University of California Press.

Rachele Ricceri, Klaas Bentein, Floris Bernard, Antoon Bronselaer, Els De Paermentier, Pieterjan De Potter, Guy De Tré, Ilse De Vos, Maxime Deforche, Kristoffel Demoen, Els Lefever, Anne-Sophie Rouckhout, and Colin Swaelens. 2023. The database of byzantine book epigrams project: Principles, challenges, opportunities. Working paper or preprint.

Helmut Schmid. 2019. Deep learning-based morphological taggers and lemmatizers for annotating historical texts. In *Proceedings of the 3rd international conference on digital access to textual cultural heritage*, pages 133–137.

Helmut Schmid and Florian Laws. 2008. Estimation of conditional probabilities with decision trees and an application to fine-grained pos tagging. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, COLING '08, page 777–784, USA. Association for Computational Linguistics.

Pranaydeep Singh, Gorik Rutten, and Els Lefever. 2021. A pilot study for BERT language modelling and morphological analysis for ancient and medieval Greek. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 128–137, Punta Cana, Dominican Republic (online). Association for Computational Linguistics.

A Appendix A

This table shows all occurrences used in the inter-annotator agreement study.

Occ. id	Tokens
17368	50
18180	33
18446	9
19604	101
20167	60
21375	43
22487	91
22734	75
23607	10
23615	12
23631	16
23632	19
25463	52
26551	66
30520	354
30844	31

Table 3: The set of epigrams used for the inter-annotator agreement study, summing up to 1,022 tokens.