An automatic model and Gold Standard for translation alignment of Ancient Greek

Tariq Yousef*, Chiara Palladino†, Farnoosh Shamsian*
Anise d'Orange Ferreira⋆, Michel Ferreira dos Reis⋆
*University of Leipzig
Augustusplatz 10, 04109 Leipzig, Germany
tariq.yosef@uni-leipzig.de, farnoosh.shamsian@uni-leipzig.de

†Furman University
3300 Poinsett Highway, 29613, Greenville SC, USA
chiara.palladino@furman.edu

⋆Universidade Estadual Paulista (UNESP)
Rod. Araraquara-Jaú Km 1 - Bairro dos Machados Machados - Araraquara/SP - CEP 14800-901, Brazil
anise.ferreira@unesp.br, michelfereis@yahoo.com.br

Abstract

This paper illustrates a workflow for the development and evaluation of automatic translation alignment models for Ancient Greek. We designed an annotation Style Guide and a gold standard for the alignment of Ancient Greek-English and Ancient Greek-Portuguese, measured inter-annotator agreement and used the resulting data set to evaluate the performance of various translation alignment models. According to our results, fine-tuned models based on XLM-Roberta are superior in performance, and we achieved impressive accuracy with the proposed training workflow even with a smaller manually aligned corpus than normally used in similar experiments.

Keywords: Translation Alignment, Gold Standard, Alignment Guidelines, Ancient Greek

1. Introduction

Word alignment is defined as the operation of comparing two or more texts in order to find correspondences between their textual units, through automated or semi-automated methods. When the texts being compared are in different languages (also called parallel texts or parallel corpora), the task is more specifically called translation alignment. The result often takes the form of a list of pairs of items, which can be larger text chunks like documents or paragraphs, but more frequently sentences and words (Kay and Röscheisen, 1993; Véronis, 2000).

Parallel corpora are used for a variety of purposes, including neural and statistical machine translation systems (DeNero and Klein, 2007), automatic bilingual lexicon extraction (Yousef, 2020), corpus linguistics (Baker, 2000), language learning (Palladino et al., 2021), and cross-lingual annotation projection (David et al., 2001; Padó and Lapata, 2009; Müller, 2017; Nicolai and Yarowsky, 2019). The challenge inherent in the field of translation alignment is that it is very difficult to achieve perfectly aligned corpora, especially at word level. In many cases, the challenge is to establish exact equivalences across languages, especially for non-literal expressions such as wordplays, allusions, rhymes, and other rhetorical devices. These issues tend to be more complex if the linguistic and cultural difference between two languages is wider, depending on genre, style, register, on the linguistic typology of source and target language, and on their degree of distance in time and space. For this reason, corpora of manually aligned texts are a much-desired resource to train and evaluate the performance of automated alignment models: in particular, some scholars developed gold standards for word alignment, which were created with manual annotation through the definition of style guides, and tested for consistency through the measurement of the performance of multiple annotators (Dagan et al., 1999; Graça et al., 2008; Lambert et al., 2005; Mareček, 2008). However, manual alignment is expensive in terms of time and resources, and requires annotation tools tailored for this purpose (Yousef and Foradi, forthcoming 2022).

To our best knowledge, most current studies on automatic translation alignment and alignment gold standards are conducted on modern languages. In this paper, we are going to illustrate a workflow for the evaluation of translation alignment models starting from Ancient Greek texts and translations into English and Brazilian Portuguese. This contribution is structured as follows: we provide a review of the related work on alignment guidelines and gold standards, and approaches to automatic word alignment. Next, we describe our work, focusing on the creation of alignment guidelines for Ancient Greek-English and Ancient Greek-Portuguese developed by domain experts and used to create reliable and high-quality word-level gold standard data sets. The gold standards were evaluated for inter-annotator agreement to ensure reliability. In the following part, we propose and fine-tune an au-

tomatic alignment model for Ancient Greek that significantly outperforms the popular statistical models such as Giza++, elforml, and fast_align; and achieves impressive results even with the absence of training data. In the closing part of the paper, we evaluate the obtained results and propose some lines of future work.

2. Related Work

The main purpose of gold standards is to help evaluate the performance of automatic alignment methods. Creating gold standards requires alignment guidelines that ensure a higher agreement among annotators, resulting in more reliable and accurate results. Therefore, most proposed works on generating gold standards were associated with developing annotation style guides. There are countless examples of the creation of alignment gold standards for pairs of modern European languages, most of them regularly including English: French-English (Melamed, 1998; Och and Ney, 2000), Dutch-English (Macken, 2010), English-Swedish (Holmqvist and Ahrenberg, 2011), Romanian-English (Mihalcea and Pedersen, 2003), Czech-English (Kruijff-Korbayová et al., 2006; Mareček, 2008), English-Spanish (Lambert et al., 2005), and English-Icelandic (Steingrímsson et al., 2021). More rarely, gold standards are produced for several language pairs, e.g. (Graça et al., 2008) developed data sets for Portuguese, English, French and Spanish. In most cases, a gold standard was created by creating new annotation guidelines or expanding existing ones.

Most alignments for this purpose were created by two different annotators, whose agreement was measured to assess the quality of the guidelines, and the reliability of the resulting gold standard. Kohen's Kappa is also another measure of the Annotator's agreement. The Inter-Annotator Agreement of the datasets mentioned above ranges from 84% to 96.5% for Spanish-French (Graça et al., 2008).

The sizes of the gold standard data sets used in the literature described above range from 100 sentences per languages pair (Graça et al., 2008), to 1500 sentences (Macken, 2010). The Europarl was the main source of parallel data for creating alignment gold standard (Graça et al., 2008; Macken, 2010; Lambert et al., 2005; Holmqvist and Ahrenberg, 2011). (Macken, 2010), however, used journalistic texts, newsletters, and medical reports to create gold standards for Dutch-English.

Most experiments employ the Sure/Possible or Sure/Possible/Null annotation schema (Holmqvist and Ahrenberg, 2011; Kruijff-Korbayová et al., 2006; Graça et al., 2008), while (Macken, 2010) employed multi-level annotation schema with Regular/Fuzzy/Null as main classes.

On Translation Alignment tools/models: The first efforts on automatic word alignment have been done by (Brown et al., 1993), who introduced statistical lexical models known as IBM models. (Och and Ney, 2003) developed Giza++ which was considered for a long time the state-of-the-art in the field of word alignment. (Dyer et al., 2013) proposed a fast and effective log-linear implementation of IBM Model2 called $fast_align$ that outperforms IBM Model 4. (Östling and Tiedemann, 2016) developed proposed $Eflomal$, an efficient and accurate word alignment model using a Bayesian model with Markov Chain Monte Carlo (MCMC) inference. The disadvantage of the statistical models that they perform poorly with the absence of training data in form of parallel sentences.

Recent studies showed the possibility of exploiting multilingual contextualized language models to create accurate alignments even with the absence of training data. (Dou and Neubig, 2021) introduced AWESOME aligner that predicts alignments from similarity matrices and proposed training objectives to fine-tune language models for better performance, (Jalili Sabet et al., 2020) proposed SimAlign that can produce high-quality word alignments without using static and contextualized embeddings. (Garg et al., 2019) proposed a supervised model that requires word alignments for training. (Alkhouli et al., 2018) introduced a model that uses the output from statistical models such as Giza++ as a supervised training for the neural model; this model tends to inherit the alignment errors from the statistical model. (Stengel-Eskin et al., 2019) introduces a discriminative neural alignment model that leverage similarity matrices of encoder-decoder representations to predict word alignments. In this paper, we use similarity matrices derived from multilingual contextualized language models such as mBert and XLM-R and employ alignment extraction approaches similar to the methods proposed by (Jalili Sabet et al., 2020; **?**). Further, we perform supervised and unsupervised fine-tuning of the models using monolingual and bilingual datasets we collected from different resources and high-quality manual alignments available on our UGARIT platform employing the training objectives introduced by (Dou and Neubig, 2021).

3. Creating a Gold Standard for Ancient Greek

The Gold Standard created for this research consisted of two data sets of manually aligned texts at word level, in Ancient Greek - English and Ancient Greek - Brazilian Portuguese respectively. The Gold Standard was created through the support of an annotation Style Guide, or Guidelines. In this section of the paper, we focus on the development of the guidelines for aligning Ancient Greek to English, which provided a model for Portuguese. The materials here described, including Guidelines and data sets, are available at `https://github.com/UgaritAlignment/Alignment-Gold-Standards`.

3.1. The Guidelines

Both the Gold Standard and the Guidelines were created by two domain experts who had worked on translation alignment with Ugarit, so no preliminary training sessions were needed. The first draft of the Style Guide was created through multiple meetings and discussions between the two experts, prior to aligning the corpus: the structure of the Guidelines was developed starting from an already existing model developed for modern languages, and considerably expanded to include various language-specific issues that the experts had encountered in their previous experience with aligning Ancient Greek texts. Then, the experts aligned a subset of the corpus to test the general consistency and feasibility of the Guidelines: during this phase, for each new issue there was a brief discussion and a preferred annotation style or an improvement in the style guide was agreed upon. Each change was incorporated in the final version of the style guide, and the alignments were revised accordingly. After the subset was completed, the experts completed the alignment without further discussions, to ensure that the consistency and efficacy of the guidelines could be appropriately tested through inter-annotator agreement.

The Guidelines for Portuguese were created in a similar fashion, as they were designed by two domain experts, who also manually aligned the corpus. These guidelines were developed from a previous draft created by a group of language experts working on a different project, and expanded substantially by taking the English Guidelines as a model.

The Guidelines consider the types of links allowed by Ugarit, which are one-to-one (1-1), many-to-many (N-N), one-to-many (1-N) and many-to-one (N-1). Links in Ugarit do not include lack of alignment (0 link): words that do not correspond are simply left unaligned. Even though Ugarit does not distinguish between possible and certain alignments, this was addressed differently in the evaluation of inter-annotator agreement (see below). The Guidelines were created specifically with the goal of creating a consistent and reliable Gold Standard to use in machine-actionable models. For this reason, the main structural problem to address was the highly inflected nature of Ancient Greek as a language, which created contrast with the translation languages. We adopted the definition of (Lambert et al., 2005): "the only valid elements in an alignment are single words and indivisible groups of words" (p.275): groups of words are linked together when the meaning of the group is distinct from that of the sequence of each word's meaning, and single words cannot be separated from the rest of the group without changing their meaning (= indivisible lexical unit). Further, we established that correspondence between lexical units had to involve as few words as possible, but as many words as necessary, with the requirement of equivalent meaning between original and translation. Basically, this meant that linguistic structures that were peculiar to Ancient Greek could be aligned as lexical units when necessary, but the general principle allowed an overall prevalence of one-to-one links, which are more useful to train computational models.

3.2. The Corpus

The corpus aligned to develop the Gold Standard included three Ancient Greek texts and translations into English and Portuguese: we included passages from the *Iliad* (2010 words), Plato's *Crito* (1829 words) and Xenophon's *Cyropedia* (1520 words). The corpus was selected according to a few guiding principles:

- It had to include enough linguistic and genre variation to be a sufficiently comprehensive example of a Gold Standard and to provide a range of issues to address in the Guidelines;
- It needed to have a consistent citation structure in both original and translation, so that both texts could be extracted already aligned at the level of a given unit, e.g. sentence, verse, or paragraph;
- The citation structure allowed selection of short text units, to facilitate the alignment and create a better training data set.

Therefore, the texts chosen provide sufficient diversity of language (Homeric to Koine Greek) and of text genre (poetry, prose, and dialogue), and could be easily aligned at the level of their specific citation unit. In the case of the *Iliad*, we used a text already aligned at sentence level with the English translation currently used in the Perseus dependency treebanks (Bamman et al., 2010). The translations we selected were mostly modern for both English and Portuguese (Murray, 1924; Burnet, 1903; Marchant, 1910; Werner, 2018). For the texts of *Crito* and *Cyropaedia*, we were kindly provided with translations in Portuguese by André Malta and Emerson Cerdas.

3.3. Further Considerations

The Guidelines revealed aspects that were consistent with other similar annotation guides for the alignment of modern languages. For example, most guidelines address punctuation, omission, phrasal construction and repetition in the same way ours do, for example: punctuation tends not to be aligned; repeated words in only one language are only aligned in the first instance; omission and ellipsis may result in lack of alignment; phrasal constructions, including idioms and proverbial expressions, are considered indivisible units and aligned N-N, and so on and so forth. However, the peculiarities of working with an ancient language often steered our approach in different directions, and especially required us to think about situations that are simply not that frequent in modern languages. Not only we had to deal with situations that are not often found in modern languages (e.g. high inflection of verbs or nouns), but we also had to consider the

very inconsistent ways in which certain parts of speech, such as articles and participles, are addressed, not only across translations, but within the same translation as well. So, while the guidelines developed for Spanish and English by (Lambert et al., 2005) (which provided the main guiding principles for our own), only included detailed guidance for 7 cases and still had a high inter-annotator agreement, our guidelines include double the number of cases (14), and go in much deeper detail to include specific variations to those situations: for instance, just our section on determiners (articles and pronouns) includes five separate situations that had to be addressed in detail for English, and six for Portuguese. The status of Ancient Greek as a dead language, where there is only a finite number of texts, also has a set of separate implications, since it is not possible to verify the accuracy of some things, which requires to make certain judgement calls in the establishment of translation pairs. For example: can a string be classified as an idiom, if it only appears in one author or one work, or even only once in the entire language corpus? How can we create consistent guidelines for phenomena, like the genitive absolute, whose semantic function is exclusively established in relation to the context where they appear?

The resulting considerations indicate that it is extremely important to understand which languages are being aligned, and to which cultural tradition they belong. For historical languages, it is simply not possible to create pre-established, easy-to-align corpora of texts with translations as close as possible to the original. Historical languages are witnesses to cultures and uses that are very distant from modern ones, and are translated in ever-changing and inconsistent ways that stretch across a time-span of centuries. As a result, guidelines will be necessarily more detailed to handle such inconsistencies and, on the other hand, there may be higher chance of internal disagreement between annotators. However, our research shows that it is possible to create them and still obtain consistency between annotators: so, it is possible to create Gold Standards, provided that strong scholarly expertise in the original language is put to use.

3.4. Alignment Results

Our alignment guidelines do not distinguish between sure and possible alignments as proposed by (Och and Ney, 2003). The alignment schema only allows sure alignments, but when combining the alignments of the two annotators, we resulted in sure and possible alignment sets for every sentence as follows:

$$S = A_1 \cap A_2 \quad , \quad P = A_1 \cup A_2$$

A_1 and A_2 are the alignments sets created by the first and second annotators, respectively. S denotes sure alignments which include all translation pairs where both annotators agree. P denotes possible alignments where the translation pairs are aligned by at least one annotator. We exported the gold standards in NAACL Format (Mihalcea and Pedersen, 2003) ....

Inter-Annotator Agreement

Inter-Annotator Agreement (IAA) is a great indicator of the reliability of the annotation guidelines and the quality of the alignment gold standards.

	Grc-Eng	Grc-Por
Sentences	275	183
Grc Tokens	5.359	3.216
Grc Types	2.347	1.587
Eng/Por Tokens	7.515	3.710
Eng/Por Types	1.634	1.355
Sure Alignments	6.240	3.028
Possible Alignments	1.423	864
IAA	86.17%	83.31%

Table 1: Inter-Annotator Agreement

We compute the IAA over the Ancient Greek-English and Ancient Greek-Portuguese data sets. Alignment agreement is considered in two cases, when both annotators align the same pair of tokens and when both annotators do not align a token. We also considered multi-word alignments (1-N, N-1, and N-N) as 1-1 pairs. For example, if the phrase "The son" is aligned to "υἱός", it is converted to two 1-1 alignments (The, υἱός) and (son, υἱός). Let A_1 and A_2 be the flattened translation pairs created by the first and second annotators, respectively, and I is the intersection between them, we calculate the IAA as follows:

$$IAA = 2 * I/(A_1 + A_2)$$

Table 1 summarizes the IAA results and provides an overview of the gold standards data sets.

4. Automatic Word Alignment Model

We use state-of-the-art automatic word alignment methods that utilize pre-trained contextualized language models to create word alignment. Further, we fine-tune a language model that can align Ancient Greek and English with a novel training strategy that combines training over monolingual and bilingual datasets, in addition to supervised training over accurate alignments provided by UGARIT. The trained model significantly outperformed all existing statistical models such as Giza++, Elfomal, and fast_align on Ancient Greek-English and Ancient Greek-Portuguese parallel texts even with the absence of any training data for Ancient Greek-Portuguese.

4.1. Algorithm

Word alignment is the process of finding word-level equivalents between the source sentence $S = (s_1, s_2, .., s_n)$ and its translation $T = (t_1, t_2, .., t_m)$ (Brown et al., 1993). The alignment process can be considered as a black box, its inputs are S and T and

its output is a set $A = \{(s_i, t_j) : s_i \in S, t_j \in T\}$ where s_i is a translation equivalent of t_j.

The core concept of recent studies (Jalili Sabet et al., 2020; Stengel-Eskin et al., 2019; Dou and Neubig, 2021) is to exploit the pre-trained multilingual contextualized language models such as mBERT (Devlin et al., 2018) and XLM-R (Conneau et al., 2019) or a fine-tuned version of them, and derive a similarity matrix based on distance/similarity measures of the contextualized word embeddings for every tokens pair. Then, the word-level alignments can be predicted by employing an extraction algorithm over this similarity matrix.

4.1.1. Similarity Matrix

Suppose S_{grc}, S_{eng} be two parallel sentences with lengths $n, m.$ and $SIM_{n \times m}$ the similarity matrix of these two sentences. Using the pre-trained contextualized word embeddings derived from multilingual transformers models, the similarity matrix can be filled as an equation 1:

$$\sum_{i}^{n} \sum_{j}^{m} SIM(i,j) = F_{sim}(t^i_{grc}, t^j_{eng}) \qquad (1)$$

Where t^i_{grc} is the embedding vector of the ith token in S_{grc}, t^j_{eng} is the embedding vector of the jth token in S_{eng}, and F_{sim} is a similarity function between the two vectors such as *Cosine Similarity*, *Dot Product*, and *Euclidean distance*.

In all our experiments, the word embeddings are extracted from 8-th layer of mBERT and XLM-R, since it has achieved the best performance according to the experiments conducted by (Jalili Sabet et al., 2020; Dou and Neubig, 2021).

Figure 1 shows an example of a similarity matrix computed using the *dot product* over the word embeddings extracted from the 8-th layer of mBERT.

4.1.2. Alignments Extraction

Once the similarity matrix is computed, alignments can be extracted by applying an extraction algorithm. (Dou and Neubig, 2021) proposed two probability thresholding-based methods to extract alignments from the similarity matrix, namely, $Softmax$ and $Entmax15$ (Peters et al., 2019). (Dou and Neubig, 2021) applies the extraction in two directions and then considers the intersection between them (Figure 1). (Dou and Neubig, 2021) applied $Softmax$ with 0.001 as threshold and $Entmax15$ with 0 as threshold. We used the same implementation and settings in our experiments.

Further, (Jalili Sabet et al., 2020) proposed three methods including $Argmax$, a baseline method , $Itermax$, an iterative method, and $Match$, a graph-based method. Jalili Sabet et al. found that $Itermax$ performs slightly better than $Argmax$, and it works better for distant languages, which was perfect for our case, since Ancient Greek and English are distant languages.

We employed the five extraction methods with their default settings in all our experiments. Section 4.4 compares and discusses in detail the performance of the various extraction methods. Figure 1 shows that the alignment is computed on subword-level. Since the task is to perform word-level alignment, converting subword-level alignments to word-level alignments is necessary. (Jalili Sabet et al., 2020; Zenkel et al., 2020; Dou and Neubig, 2021) employed the heuristic principle *"two words are aligned if any of their subwords are aligned"*, which we also followed.

4.2. Training Process and Objectives

The experiments we conducted on the pre-trained mBERT and XLM-R (Zero-Shot) showed significantly poor performance on both Ancient Greek-English and Ancient Greek-Portuguese data sets. Therefore, it was necessary to train and fine-tune those models aiming for better performance. Due to the availability of parallel sentences and in order to obtain the best outcome from the training process, we employed the training objectives proposed by (Dou and Neubig, 2021):

- **Masked Language Modeling (MLM):** following (Gururangan et al., 2020), Dou and Neubig propose to fine-tune language models with MLM on the source and target sentences.
- **Translation Language Modeling (TLM):** this objective concatenates the source and target sentences and perform MLM on the resulting string, which allows the model to align the source and target representations.
- **Self-training Objective (SO):** this objective enhances the symmetricity between forward and backward alignments and encourages words aligned in the first alignment pass to have closer embeddings.
- **Parallel Sentence Identification (PSI):** this objective encourages parallel sentences to be more similar than random unaligned sentences, which enables the overall alignments of embeddings to be closer together.

We conducted the following experiments:

Experiment 1

In this experiment we performed unsupervised fine-tuning of the mBert and XLM-R using 32500 Ancient Greek-English parallel sentences with the training objectives MLM, TLM, SO, PSI. The parallel sentences used in this experiment are taken from Perseus Digital Library[1] (Iliad, Odyssey, Xenophon, New Testament).

[1] https://github.com/PerseusDL/canonical-greekLit

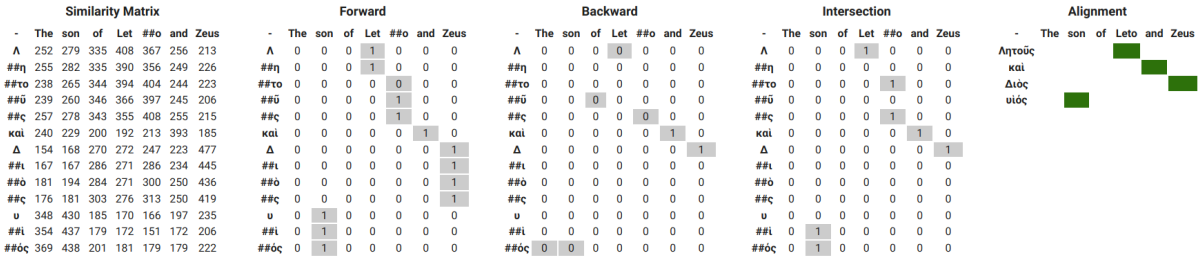


Figure 1: .....

Experiment	Input Models	Epochs	Training Objectives	Languages	Data Size	Source
EX 1	mBERT, XLM-R	1	MLM, SO, TLM, PSI	GRC-ENG	32.500 parallel sentences	Perseus
EX 2	EX1 Fine tuned models	1	MLM, SO TLM, PSI	GRC-LAT	8.000 parallel sentences	DFHG
EX 3	EX2 Fine tuned models	5	MLM	GRC Monolingual	12 Millions Tokens	Perseus, First1kGreek, TreeBanking
EX 4	EX3 Fine tuned models	5	MLM, SO TLM, PSI	GRC-ENG GRC-LAT GRC-KAT	45.000 parallel sentences	Perseus, DFHG, UGARIT
EX 5	EX3 Fine tuned models	5	MLM, SO TLM, PSI			
EX 6	EX5 Fine tuned models	15	SO	Mixed dataset	2200 parallel sentences, 100k Translation Pairs	UGARIT

Table 2: .....

Experiment 2

In this experiment, we also performed unsupervised training of the fine-tuned models of Experiment 1. We used 8000 Ancient Greek-Latin parallel Fragments with the same training objectives in the previous experiment. The parallel fragments are taken from the Digital Fragmenta Historicorum Graecorum project [2].

Experiment 3

The existing bilingual contextualized language models mBERT and XLM-R are not trained on ancient Greek texts but on modern Greek, which is very different. So it was necessary to train them on monolingual Ancient Greek texts. In this experiment, We trained the fine-tuned models we obtained after Experiment 2 on 12 million Ancient Greek tokens with Masked Language Model training objective. The training dataset is extracted from Perseus Digital Library, the first thousand years of Greek project[3], and the PROIEL, PERSEUS[4], and Gorman[5] treebanking datasets.

[2]https://www.dfhg-project.org/

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

[4]universaldependencies.org

[5]https://vgorman1.github.io/

Experiment 4

This experiment and the next one aim to inspect the impact of training the models on monolingual texts; therefore, the two experiments use the same training data, but they differ by the training model. This experiment trains the model obtained after Experiment 2 on 45000 parallel texts (We combined the datasets in *Experiments 1 and 2* with 4000 further parallel sentences taken from UGARIT database. For training, we used MLM, TLM, SO, PSI training objectives. The texts are in different languages mainly (Ancient Greek-English, Ancient Greek-Latin, and Ancient Greek-Georgian)

Experiment 5

In this experiment, we train the model obtained after *Experiments 3* which is trained on monolingual Ancient Greek texts. For training, we use the same training dataset used in *Experiments 4* with the training objectives MLM, TLM, SO, PSI.

Experiment 6

In this experiment, we perform supervised training for the fine-tuned model obtained after *Experiment 5* using word-level manually aligned dataset provided by UGARIT. The alignments are accurate and clean since they are done by Scholar, teachers, and Experts. The datasts consists of 2265 parallel texts and almost 100k

translation pairs.

4.3. Evaluation

4.3.1. Baseline Models

We compare our model to three popular statistical word alignment models, namely, Giza++ (Och and Ney, 2003), fast_align (Dyer et al., 2013) , elfomal (Östling and Tiedemann, 2016). All these models require training data in the form of parallel sentences. We trained them on the 35000 Ancient Greek-English parallel sentences and 275 gold standard sentences and evaluated their performance on our produced gold standards. Table 3 shows the results of our evaluation. The poor performance on the Ancient Greek Portuguese dataset is because of the absence of the training data; the models are trained only on 183 sentences of the gold standard dataset.

4.3.2. Evaluation Metrics

Similar to (Och and Ney, 2003), we evaluate the performance of the alignment model against the gold standards, by employing $Precision$, $Recall$, $F1$, and Alignment Error Rate (AER) , which can be computed as in equations 2.

$$Precision = \frac{|A \cap P|}{|A|} \quad , \quad Recall = \frac{|A \cap S|}{|S|}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{2}$$

$$AER = 1 - \frac{|A \cap P| + |A \cap S|}{|A| + |S|}$$

Where A indicates the alignments set predicted by the model, P and S indicate respectively the $Possible$ and $Sure$ alignment sets in the gold standards, and $|.|$ denotes the length of the set.

4.4. Results and Conclusions

The results reported in tables 3 and 4 undoubtedly shows the superiority of the fine-tuned models over the statistical models (Giza++, elfomal, fast_align). Furthermore, the last two experiment (*Experiments 5 and 6*) show that the alignments derived from fine-tuned XLM-R models are superior to those derived from mBERT fine-tuned models for Ancient Greek/English and Ancient Greek/Portuguese datasets.

The results of *Experiment 2* shows that training the model on Ancient Greek-Latin fragments enhanced the alignment results and decreased the AER by 0.79%-0.9% for XLM-R fine-tuned model and by 2.06%-2.42% for the fine-tuned mBert model.

The results indicate that training the model on monolingual data (EX3) decreased the $F1$ and increased the AER on both models. As we mentioned before, *Experiments 4 and 5* use the same training datasets and the same number of epochs, but we can see the great difference in the results, where all *Experiment 5* evaluation metrics are superior to *Experiment 4* in both model, which undoubtedly shows the impact of training the models on monolingual datasets.

Table 4 shows that the proposed training strategy achieved impressive results on Ancient Greek/Portuguese parallel texts in the complete absence of any training data.

Regarding the alignment extraction approaches, table 3 shows that $Itermax$ achieved the lowest AER and the highest $F1$ in Experiments 6 and 7 on both models. Whereas $Argmax$ achieved the highest $Precision$ in the same mentioned experiments. Also $Softmax$ achieved better AER, $F1$, $Recall$ than $Entmax15$ in all experiments, while $Entmax15$ is always superior regarding $Precision$. The results in table 4 look different, the $Softmax$ achieved the best AER, $F1$ in experiment 6 on the fine-tuned mBert model, whereas $Itermax$ achieved the best AER, $F1$ in the same experiment but on fine-tuned XML-R model.

The tables also show that fine-tuning the multilingual models on monolingual dataset played a key role of enhancing the model performance. Further, the performance of the model can be enhanced by performing supervised and unsupervised fine-tuning even on languages pairs that are different from the target language pairs.

Moreover, If monolingual or bilingual texts are available, we recommend to fine-tune the model on monolingual data with Masked Language Modeling training objective first , then perform supervised fine-tuning with the desired training objectives.

The tables show also the great impact of supervised training with word-level manual alignments which decreased the AER from 31.85% to 18.27% on fine-tuned XML-R model and from 37.69% to 32.47% on fine-tuned mBert model. Which leads us to a conclusion that fine-tuned XMl-R model are more sensitive to the supervised fine-tuning that fine-tuned mBert models.

5. Future Work

Our experiments showed the great impact of fine-tuning pre-trained multilingual contextualized models on both mono- and bilingual training datasets on the accuracy of automatic word alignment models. As future work we plan to train the model on more data:

- *Monolingual Training:* We intend to train the model on more monolingual Ancient Greek texts, and for more epochs, the current model is trained on 12 million tokens and for only five epochs.

- *Unsupervised Multi-lingual Training:* We plan to thoroughly inspect Perseus Digital Library to collect more Ancient Greek parallel texts that are accurately aligned at some level (paragraph or sentence) and train the model on the new texts for more epochs.

	Experemint	Extraction Method	mBert			
			Precision	Recall	F1	AER
Statistical Models	Giza++		37.25%	29.26%	32.78%	67.01%
	fast_align		37.37%	35.64%	36.48%	63.47%
	eflomal		47.17%	42.93%	44.95%	54.95%
mBERT	Zero Shot	Softmax	37.14%	21.09%	26.90%	72.70%
		Entmax15	42.58%	17.34%	24.64%	74.94%
	EX1	Softmax	52.98%	38.21%	44.40%	55.28%
		Entmax15	56.96%	35.50%	43.74%	55.84%
	EX2	Softmax	55.89%	40.24%	46.79%	52.86%
		Entmax15	59.84%	37.04%	45.76%	53.78%
	EX3	Softmax	54.03%	39.68%	45.76%	53.94%
		Entmax15	60.35%	35.29%	44.54%	54.99%
	EX4	Softmax	65.06%	48.08%	55.30%	44.33%
		Entmax15	68.23%	46.01%	54.96%	44.58%
	EX5	Softmax	74.12%	54.01%	62.49%	37.11%
		Entmax15	77.22%	51.68%	61.92%	37.62%
		Match	69.42%	53.56%	60.47%	39.19%
		Argmax	80.12%	48.85%	60.69%	38.78%
		Itermax	71.20%	54.84%	61.96%	37.69%
	EX6	Softmax	80.29%	**56.36%**	66.23%	33.28%
		Entmax15	83.53%	53.25%	65.04%	34.41%
		Match	71.10%	55.29%	62.21%	37.47%
		Argmax	**85.63%**	49.21%	62.50%	36.91%
		Itermax	76.83%	59.68%	**67.18%**	**32.47%**
XLM-R	Zero Shot	Softmax	37.59%	11.84%	18.01%	81.80%
		Entmax15	46.74%	8.67%	14.63%	85.20%
	EX1	Softmax	54.61%	28.21%	37.20%	62.46%
		Entmax15	65.35%	22.76%	33.76%	65.86%
	EX2	Softmax	55.62%	28.97%	38.10%	61.56%
		Entmax15	65.14%	23.49%	34.53%	65.07%
	EX3	Softmax	53.58%	24.21%	33.35%	66.33%
		Entmax15	61.99%	18.54%	28.54%	71.11%
	EX4	Softmax	65.22%	36.39%	46.72%	52.88%
		Entmax15	73.42%	30.34%	42.94%	56.57%
	EX5	Softmax	76.41%	53.88%	63.20%	36.45%
		Entmax15	82.36%	48.32%	60.91%	38.66%
		Match	62.35%	72.63%	67.10%	33.14%
		Argmax	84.52%	51.68%	64.14%	35.42%
		Itermax	74.37%	62.52%	67.93%	31.85%
	EX6	Softmax	90.85%	67.48%	77.44%	22.12%
		Entmax15	92.89%	63.62%	75.52%	24.01%
		Match	77.02%	**84.62%**	80.64%	19.57%
		Argmax	**93.04%**	62.23%	74.58%	24.94%
		Itermax	87.44%	76.30%	**81.49%**	**18.27%**

Table 3: Evaluation results on Ancient Greek-English gold standards

- *Supervised Multi-lingual Training:* Since we were keen to train a high-quality alignment model, the current model is trained on the manual alignments of only 10 trusted users (experts, scholars, and teachers). We ignored all other alignments done by students or by users that we don't know. We will carefully check the existing alignments where Ancient Greek is one of the aligned languages, select the accurate alignments and include them in the training data.

The great performance achieved by fine-tuning the model on word alignment task has encouraged us to test the model on other downstream tasks such as named-entity recognition and Part-Of-Speech tagging of Ancient Greek texts.

Finally, we intend to create more gold standards for other language pairs such as Ancient Greek/Latin, Ancient Greek/Italian, and Ancient Greek/Persian and test the model performance on the new datasets.

	Experemint	Extraction Method	mBert			
			Precision	Recall	F1	AER
Statistical Models	Giza++		25.59%	24.60%	0.2509	74.88%
	fast_align		25.62%	30.14%	27.70%	72.47%
	eflomal		34.84%	35.59%	35.21%	64.81%
mBERT	Zero Shot	Softmax	30.08%	26.66%	28.27%	71.66%
		Entmax15	33.65%	22.67%	27.09%	72.73%
	EX6	Softmax	63.84%	**61.27%**	**62.53%**	**37.40%**
		Entmax15	65.49%	57.41%	61.18%	38.61%
		Match	53.61%	59.47%	56.39%	43.78%
		Argmax	**66.21%**	51.87%	58.17%	41.45%
		Itermax	58.68%	61.19%	59.91%	40.17%
XLM-R	Zero Shot	softmax	21.68%	13.53%	16.66%	83.16%
		entmax15	23.55%	9.27%	13.30%	86.44%
	EX6	Softmax	76.11%	75.61%	75.86%	24.13%
		Entmax15	**77.45%**	72.69%	74.99%	24.89%
		Match	58.79%	**86.17%**	69.89%	31.01%
		Argmax	77.25%	71.10%	74.05%	25.81%
		Itermax	72.22%	81.02%	**76.37%**	23.91%

Table 4: Evaluation results on Ancient Greek-Portuguese gold standards

6. Acknowledgements

We are grateful to the Ugarit community of experts and annotators, who provided the necessary data and support for this research. Moreover, we thank scholars André Malta and Emerson Cerdas for providing the Portuguese translations of the texts used in our corpus; and Gregory Crane, for the guidance and support of the Ugarit project.

7. Bibliographical References

Alkhouli, T., Bretschner, G., and Ney, H. (2018). On the alignment problem in multi-head attention-based neural machine translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 177–185, Brussels, Belgium, October. Association for Computational Linguistics.

Baker, M. (2000). Towards a Methodology for Investigating the Style of a Literary Translator.

Bamman, D., Mambrini, F., and Crane, G. (2010). An ownership model of annotation: The ancient greek dependency treebank. 12.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

Burnet, J. (1903). *Platonis opera. T. 3: Tetralogias V - VII continens*. Scriptorum classicorum bibliotheca Oxoniensis. Clarendon, Oxonii, nachdr. edition.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Dagan, I., Church, K., and Gale, W. (1999). Robust Bilingual Word Alignment for Machine Aided Translation. In Susan Armstrong, et al., editors, *Natural Language Processing Using Very Large Corpora*, Text, Speech and Language Technology, pages 209–224. Springer Netherlands, Dordrecht.

David, Y., Grace, N., Richard, W., et al. (2001). Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the First International Conference on Human Language Technology Research*, pages 1–8.

DeNero, J. and Klein, D. (2007). Tailoring word alignments to syntactic machine translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 17–24.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Dou, Z.-Y. and Neubig, G. (2021). Word alignment by fine-tuning embeddings on parallel corpora. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2112–2128, Online, April. Association for Computational Linguistics.

Dyer, C., Chahuneau, V., and Smith, N. A. (2013). A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia, June. Association for Computational Linguistics.

Garg, S., Peitz, S., Nallasamy, U., and Paulik, M. (2019). Jointly learning to align and translate with transformer models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International*

Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 4453–4462, Hong Kong, China, November. Association for Computational Linguistics.

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., and Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online, July. Association for Computational Linguistics.

Jalili Sabet, M., Dufter, P., Yvon, F., and Schütze, H. (2020). SimAlign: High quality word alignments without parallel training data using static and contextualized embeddings. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1627–1643, Online, November. Association for Computational Linguistics.

Kay, M. and Röscheisen, M. (1993). Text-translation Alignment. *Comput. Linguist.*, 19(1):121–142, March.

Marchant, E. C. (1910). *Xenophontis Opera omnia*. Clarendon Press. OCLC: 802674413.

Melamed, I. D. (1998). Manual annotation of translational equivalence: The blinker project. *arXiv preprint cmp-lg/9805005*.

Müller, M. (2017). Treatment of markup in statistical machine translation. Association of Computational Linguistics.

Murray, A. (1924). *The Iliad. With an English translation by A.T. Murray*. William Heinemann ; G.P. Putnam's Sons London (England) : New York (New York).

Nicolai, G. and Yarowsky, D. (2019). Learning morphosyntactic analyzers from the Bible via iterative annotation projection across 26 languages. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1765–1774, Florence, Italy, July. Association for Computational Linguistics.

Och, F. J. and Ney, H. (2000). Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, ACL '00, page 440–447, USA. Association for Computational Linguistics.

Och, F. J. and Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

Östling, R. and Tiedemann, J. (2016). Efficient word alignment with Markov Chain Monte Carlo. *Prague Bulletin of Mathematical Linguistics*, 106:125–146, October.

Padó, S. and Lapata, M. (2009). Cross-lingual annotation projection for semantic roles. *Journal of Artificial Intelligence Research*, 36:307–340.

Palladino, C., Foradi, M., and Yousef, T. (2021). Translation alignment for historical language learning: a case study. *Digital Humanities Quarterly*, 15(3).

Peters, B., Niculae, V., and Martins, A. F. (2019). Sparse sequence-to-sequence models. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1504–1519.

Steingrímsson, S., Loftsson, H., and Way, A. (2021). CombAlign: a tool for obtaining high-quality word alignments. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 64–73, Reykjavik, Iceland (Online), May 31–2 June. Linköping University Electronic Press, Sweden.

Stengel-Eskin, E., Su, T.-r., Post, M., and Van Durme, B. (2019). A discriminative neural model for cross-lingual word alignment. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 910–920, Hong Kong, China, November. Association for Computational Linguistics.

Jean Véronis, editor. (2000). *Parallel Text Processing: Alignment and Use of Translation Corpora*. Text, Speech and Language Technology. Springer Netherlands, Dordrecht-Boston-London.

Werner, C. (2018). *Homero, Ilíada*. Ubu Editora, 1ª edição edition, November.

Yousef, Tariq, P. C. S. F. and Foradi, M. (forthcoming 2022). Translation alignment with ugarit. *Information*.

Yousef, T. (2020). Ugarit: Translation alignment visualization.

Zenkel, T., Wuebker, J., and DeNero, J. (2020). End-to-end neural word alignment outperforms GIZA++. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1605–1617, Online, July. Association for Computational Linguistics.

8. Language Resource References

Graça, J., Pardal, J. P., Coheur, L., and Caseiro, D. (2008). Building a golden collection of parallel multi-language word alignment. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).

Holmqvist, M. and Ahrenberg, L. (2011). A gold standard for English-Swedish word alignment. In *Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA 2011)*, pages 106–113, Riga, Latvia, May. Northern European Association for Language Technology (NEALT).

Kruijff-Korbayová, I., Chvátalová, K., and Postolache, O. (2006). Annotation guidelines for czech-english word alignment. In *LREC*, pages 1256–1261. Citeseer.

Lambert, P., DE GISPERT, A., BANCHS, R., and MARINO, J. B. (2005). Guidelines for word alignment evaluation and manual alignment. *Language resources and evaluation*, 39(4):267–285.

Macken, L. (2010). An annotation scheme and gold standard for dutch-english word alignment. In *7th conference on International Language Resources and Evaluation (LREC 2010)*, pages 3369–3374. European Language Resources Association (ELRA).

Mareček, D. (2008). Automatic alignment of tectogrammatical trees from czech-english parallel corpus. Master's thesis, Charles University, MFF UK.

Mihalcea, R. and Pedersen, T. (2003). An evaluation exercise for word alignment. In *Proceedings of the HLT-NAACL 2003 Workshop on Building and using parallel texts: data driven machine translation and beyond*, pages 1–10.

Forward

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.43	0.0	0.0	0.57
##ι	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	1.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ός	0.0	0.0	0.0	0.0	0.0	0.0	1.0

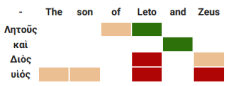
Backward

-	The	son	of	Let	##o	and	Zeus
Λ	1.0	0.02	0.0	0.26	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.98	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.69	0.0	0.0	0.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ι	0.0	0.0	0.0	0.05	0.0	0.0	0.0
##ός	0.0	0.0	1.0	0.0	0.0	0.0	1.0

Intersection

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ι	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ός	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Alignment



Pre: 40.0%, **Recall:** 33.0%, **F1:** 36.0%, **AER:** 36.0%

Zero-shot mBert (Softmax)

Forward

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
η	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ς	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.01	0.0	0.0	0.99
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
_	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.5
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
ἰ	0.0	0.62	0.0	0.0	0.0	0.0	0.02	0.36
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

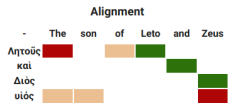
Backward

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	0.56	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	0.53	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	0.03	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_	0.0	1.0	0.14	0.44	0.47	0.0	0.97	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ἰ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.86	0.0	0.0	0.0	0.0	1.0

Intersection

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ἰ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Alignment



Pre: 60.0%, **Recall:** 50.0%, **F1:** 55.0%, **AER:** 55.0%

Zero-shot XLM-R (Softmax)

Forward

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.43	0.0	0.0	0.57
##ι	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	1.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ός	0.0	0.0	0.0	0.0	0.0	0.0	1.0

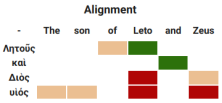
Backward

-	The	son	of	Let	##o	and	Zeus
Λ	1.0	0.02	0.0	0.26	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.98	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.69	0.0	0.0	0.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ι	0.0	0.0	0.0	0.05	0.0	0.0	0.0
##ός	0.0	0.0	1.0	0.0	0.0	0.0	1.0

Intersection

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ι	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ός	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Alignment



Pre: 40.0%, **Recall:** 33.0%, **F1:** 36.0%, **AER:** 36.0%

Zero-shot mBert (entmax15)

Forward

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
η	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ς	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.01	0.0	0.0	0.99
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
_	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.5
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
ἰ	0.0	0.62	0.0	0.0	0.0	0.0	0.02	0.36
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

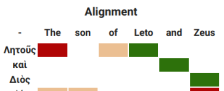
Backward

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	0.56	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	0.53	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	0.03	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_	0.0	1.0	0.14	0.44	0.47	0.0	0.97	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ἰ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.86	0.0	0.0	0.0	0.0	1.0

Intersection

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ἰ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Alignment



Pre: 60.0%, **Recall:** 50.0%, **F1:** 55.0%, **AER:** 55.0%

Zero-shot XLM-R (entmax15)

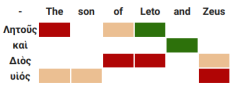
Similarity Matrix

-	The	son	of	Let	##o	and	Zeus
Λ	0.79	0.74	0.73	0.79	0.76	0.73	0.77
##η	0.69	0.69	0.68	0.74	0.73	0.68	0.71
##το	0.71	0.68	0.71	0.74	0.76	0.69	0.71
##ῦ	0.7	0.68	0.68	0.71	0.72	0.68	0.71
##ς	0.74	0.74	0.74	0.74	0.78	0.72	0.78
και	0.76	0.73	0.73	0.72	0.72	0.81	0.76
Δ	0.75	0.71	0.71	0.78	0.74	0.72	0.79
##ι	0.68	0.66	0.66	0.71	0.7	0.65	0.68
##ὸ	0.69	0.67	0.67	0.67	0.69	0.67	0.71
##ς	0.73	0.7	0.73	0.71	0.73	0.72	0.77
υ	0.74	0.72	0.72	0.74	0.74	0.73	0.77
##ι	0.71	0.71	0.71	0.78	0.73	0.7	0.77
##ός	0.75	0.73	0.74	0.72	0.77	0.73	0.81

IterMax

-	The	son	of	Let	##o	and	Zeus
Λ	1.0	0.0	0.0	0.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	1.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ός	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Alignment



Pre: 33.0%, **Recall:** 33.0%, **F1:** 33.0%, **AER:** 33.0%

Zero-shot mBert (IterMax)

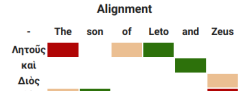
Similarity Matrix

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.84	0.86	0.84	0.9	0.86	0.85	0.87	0.86
η	0.88	0.81	0.83	0.85	0.83	0.81	0.82	0.81
το	0.88	0.84	0.85	0.83	0.88	0.84	0.83	0.84
ῦ	0.78	0.81	0.81	0.82	0.82	0.8	0.81	0.82
ς	0.89	0.84	0.85	0.82	0.84	0.85	0.81	0.86
_και	0.85	0.86	0.84	0.83	0.84	0.89	0.83	0.86
_Δι	0.83	0.86	0.83	0.86	0.86	0.85	0.88	0.86
ὸ	0.79	0.81	0.8	0.8	0.82	0.81	0.82	0.82
ς	0.87	0.82	0.84	0.8	0.82	0.83	0.8	0.85
_	0.86	0.88	0.87	0.86	0.87	0.89	0.88	0.89
υ	0.82	0.85	0.83	0.84	0.84	0.85	0.84	0.87
ἰ	0.78	0.83	0.8	0.81	0.8	0.81	0.83	0.83
ός	0.83	0.86	0.86	0.82	0.86	0.86	0.83	0.89

IterMax

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ἰ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Alignment



Pre: 60.0%, **Recall:** 50.0%, **F1:** 55.0%, **AER:** 55.0%

Zero-shot XLM-R (IterMax)

Forward

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	1.0
υ	0.0	1.0	0.0	0.0	0.0	0.0	0.0
##ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0
##ός	0.0	1.0	0.0	0.0	0.0	0.0	0.0

Backward

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.1	0.0	0.02	0.0	0.0
##ῦ	0.0	0.0	0.86	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.04	0.0	0.98	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ί	0.0	0.22	0.0	0.0	0.0	0.0	0.0
##ός	1.0	0.78	0.0	0.0	0.0	0.0	0.0

Intersection

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0
##ός	0.0	1.0	0.0	0.0	0.0	0.0	0.0

Alignment

Pre: 100.0%, **Recall:** 67.0%, **F1:** 80.0%, **AER:** 80.0%

Fine-tuned mBert (Softmax)

Forward

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
_	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0

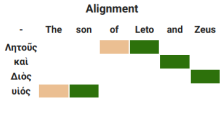
Backward

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
η	0.02	0.0	0.0	1.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
ς	0.98	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.06
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.94
_	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.08	0.0	0.0	0.0	0.0	0.0	0.0
ί	0.0	0.92	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Intersection

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
_	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Alignment



Pre: 100.0%, **Recall:** 67.0%, **F1:** 80.0%, **AER:** 80.0%

Fine-tuned XLM-R (Softmax)

Forward

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	1.0
υ	0.0	1.0	0.0	0.0	0.0	0.0	0.0
##ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0
##ός	0.0	1.0	0.0	0.0	0.0	0.0	0.0

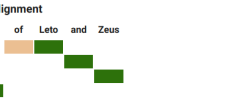
Backward

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.1	0.0	0.02	0.0	0.0
##ῦ	0.0	0.0	0.86	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.04	0.0	0.98	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ί	0.0	0.22	0.0	0.0	0.0	0.0	0.0
##ός	1.0	0.78	0.0	0.0	0.0	0.0	0.0

Intersection

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	1.0	0.0	0.0
##ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0
##ός	0.0	1.0	0.0	0.0	0.0	0.0	0.0

Alignment



Pre: 100.0%, **Recall:** 67.0%, **F1:** 80.0%, **AER:** 80.0%

Fine-tuned mBert (entmax15)

Forward

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
_	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0

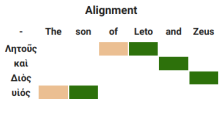
Backward

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
η	0.02	0.0	0.0	1.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
ς	0.98	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.06
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.94
_	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.08	0.0	0.0	0.0	0.0	0.0	0.0
ί	0.0	0.92	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Intersection

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
_	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Alignment



Pre: 100.0%, **Recall:** 67.0%, **F1:** 80.0%, **AER:** 80.0%

Fine-tuned XLM-R (entmax15)

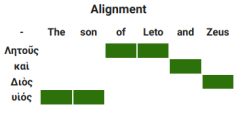
Similarity Matrix

-	The	son	of	Let	##o	and	Zeus
Λ	0.72	0.74	0.79	0.84	0.82	0.72	0.67
##η	0.72	0.74	0.79	0.83	0.81	0.72	0.69
##το	0.71	0.73	0.79	0.83	0.85	0.71	0.68
##ῦ	0.71	0.72	0.8	0.81	0.85	0.72	0.67
##ς	0.73	0.74	0.8	0.81	0.87	0.73	0.68
και	0.71	0.7	0.67	0.66	0.69	0.85	0.65
Δ	0.63	0.64	0.72	0.72	0.71	0.69	0.87
##ι	0.64	0.64	0.74	0.72	0.74	0.7	0.85
##ὸ	0.65	0.66	0.74	0.72	0.75	0.71	0.85
##ς	0.65	0.65	0.76	0.73	0.77	0.72	0.84
υ	0.8	0.87	0.66	0.64	0.64	0.67	0.69
##ί	0.8	0.87	0.65	0.64	0.63	0.65	0.67
##ός	0.82	0.88	0.67	0.65	0.66	0.66	0.68

IterMax

-	The	son	of	Let	##o	and	Zeus
Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0
##η	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##το	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ῦ	0.0	0.0	1.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	1.0	0.0	0.0
και	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Δ	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ι	0.0	0.0	0.0	0.0	0.0	0.0	1.0
##ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0
##ί	1.0	0.0	0.0	0.0	0.0	0.0	0.0
##ός	0.0	1.0	0.0	0.0	0.0	0.0	0.0

Alignment



Pre: 100.0%, **Recall:** 100.0%, **F1:** 100.0%, **AER:** 100.0%

Fine-tuned mBert (IterMax)

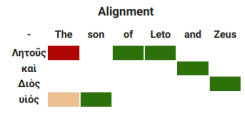
Similarity Matrix

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.74	0.73	0.74	0.85	0.8	0.64	0.64	0.62
η	0.75	0.72	0.73	0.85	0.83	0.64	0.62	0.61
το	0.73	0.73	0.74	0.82	0.86	0.66	0.62	0.62
ῦ	0.74	0.75	0.78	0.81	0.85	0.68	0.64	0.65
ς	0.77	0.76	0.76	0.79	0.83	0.67	0.63	0.65
_και	0.64	0.65	0.64	0.65	0.65	0.88	0.66	0.67
_Δι	0.61	0.64	0.68	0.68	0.68	0.72	0.87	0.85
ὸ	0.62	0.65	0.7	0.67	0.69	0.72	0.86	0.87
ς	0.62	0.64	0.7	0.66	0.68	0.71	0.85	0.89
_	0.76	0.81	0.62	0.6	0.6	0.67	0.67	0.66
υ	0.74	0.82	0.61	0.59	0.59	0.65	0.66	0.65
ί	0.74	0.82	0.61	0.59	0.59	0.65	0.67	0.65
ός	0.75	0.8	0.62	0.6	0.61	0.67	0.68	0.68

IterMax

-	_The	_son	_of	_Le	to	_and	_Ze	us
_Λ	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
η	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
το	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
ῦ	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
ς	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
_και	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
_Δι	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
ὸ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
ς	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
_	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
υ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ί	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
ός	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Alignment



Pre: 83.0%, **Recall:** 83.0%, **F1:** 83.0%, **AER:** 83.0%

Fine-tuned XLM-R (IterMax)