Graecia capta ferum victorem cepit Detecting Latin Allusions to Ancient Greek Literature

Frederick Riemenschneider **Anette Frank**
Department of Computational Linguistic
Heidelberg University
69120 Heidelberg, Germany
{riemenschneider|frank}@cl.uni-heidelberg.de

Abstract

Intertextual allusions hold a pivotal role in Classical Philology, with Latin authors frequently referencing Ancient Greek texts. Until now, the automatic identification of these intertextual references has been constrained to monolingual approaches, seeking parallels solely within Latin or Greek texts. In this study, we introduce SPHILBERTA, a trilingual Sentence-ROBERTA model tailored for Classical Philology, which excels at cross-lingual semantic comprehension and identification of identical sentences across Ancient Greek, Latin, and English. We generate new training data by automatically translating English texts into Ancient Greek. Further, we present a case study, demonstrating SPHILBERTA's capability to facilitate automated detection of intertextual parallels. Our models and resources are available at https://github.com/Heidelberg-NLP/ancient-language-models.

1 Introduction

The study of intertextuality and allusions to literary sources has a longstanding tradition in Classical Philology, highlighting complex interconnections between different literary works. During the 1960s, the concept of intertextuality was shaped by a comprehensive theoretical framework developed by scholars such as Julia Kristeva, Ferdinand de Saussure, and Michail Bakhtin. The term "intertextuality" itself was introduced by Kristeva during this pivotal era (Alfaro, 1996; Bendlin, 2006; Kristeva, 1986; Orr, 2003).

Intertextuality proves particularly crucial when examining Roman literature's relationship with Ancient Greek texts. Many Latin authors consciously mirrored elements of Greek classics, making intertextuality an essential concept for understanding this cultural literary exchange.[1]

The importance of intertextuality, especially given the considerable attention it has received, is beyond dispute. While there exists a plethora of theoretical work exploring specific forms of intertextuality, our focus in this work is on the general occurrence of textual resemblances, specifically within Latin and Greek texts.

Traditionally, the identification of such parallels has largely relied on scholars' close reading. However, recent years have seen the development of statistical NLP tools – driven especially by the Tesserae project (Coffee et al., 2012; Forstall et al., 2014) at the forefront of this movement – that are able to automatically uncover a considerable number of textual parallels. These approaches, however, typically rely on string-level parallels and are grounded in carefully designed rules and scoring functions. Notably, these systems are generally restricted to detecting parallels in the same language, as they rely on identifying identical tokens or stems.

Recently, the breakthrough in self-supervised training of powerful pre-trained language models (PLMs) has also led to a surge of diverse PLMs for Classical Philology (Bamman and Burns, 2020; Yamshchikov et al., 2022; Mercelis and Keersmaekers, 2022; Singh et al., 2021; Riemenschneider and Frank, 2023). In fact, two recent case studies in Bamman and Burns (2020) and Burns (2023) have shown that contextualized embeddings produced by such models can indeed identify texts bearing similar content. While a rigorous quantitative evaluation of these findings still remains to be conducted, the perceived potential of using these models for finding intertextual relations is clearly sparking widespread interest.

However, research into modern language analysis tasks has demonstrated that sentence embeddings derived solely from standalone BERT- or ROBERTA-based models generate suboptimal and inefficient embeddings. This insight led to

[1] Cf. Hutchinson (2013): "How Latin literature relates to Greek literature is one of the most fundamental questions for Latin literature, and for the reception of Greek."

Proceedings of the Ancient Language Processing Workshop associated with RANLP-2023, pages 30–38,
held in Varna Bulgaria, Sept 8, 2023.
https://doi.org/10.26615/978-954-452-087-8_004

the creation of Sentence-BERT (SBERT) models (Reimers and Gurevych, 2019).

Among the latest language models introduced in the field of Classical Philology is PHILBERTA (Riemenschneider and Frank, 2023), a ROBERTA-based model pre-trained on Ancient Greek, Latin, and English language data. Building upon this model, we present SPHILBERTA, a model tailored to the discovery of intertextual parallels across Latin, Ancient Greek, and English texts.

In this work, our goal is to move away from systems relying on hand-crafted rules, and instead to employ state-of-the-art tools for identifying intertextual relations that are easy to adapt to a wide variety of languages from Classical Philology and beyond. Most importantly, we probe the feasibility of uncovering intertextual parallels *across languages*, an area that has been largely neglected in the automatic identification of intertextual allusions until this point. This novel capability will considerably enlargen the space for new findings, by being able to compare texts directly across languages.

We show that SPHILBERTA is proficient in recognizing direct translations of sentences in Ancient Greek, Latin, and English, thereby demonstrating comprehensive cross-lingual competence. Applying our model directly to texts of philological significance not only underlines its practical applicability but also highlights areas for improvement, suggesting promising avenues for future exploration.

In summary, our contributions are as follows:

i) We introduce SPHILBERTA, a multilingual sentence transformer for Latin, Ancient Greek, and English. To our knowledge, we are the first to apply this type of model to automatically detect passages of potential *cross-lingual* allusions in Latin texts.
ii) To alleviate the scarcity of parallel sentence pairs for training SPHILBERTA, we augment the available resources by automatically translating English texts to Ancient Greek using an existing multilingual T5 model pre-trained on Ancient Greek, Latin, and English data.
iii) We conduct experiments on retrieving translations or similar sentences from textual passages in foreign-language texts, using cross-lingual SPHILBERTA sentence embeddings.
iv) Our experiments demonstrate that SPHILBERTA is able to detect translations with high accuracy and that data augmentation significantly enhances the performance of the system for Ancient Greek. While finding textual allusions still requires philological expertise, we present cases where the model identifies passages linked to known allusive texts.

2 Related Work

Detecting Intertextual Allusions. Initiated in 2008, the Tesserae project (Coffee et al., 2012; Forstall et al., 2014) has been instrumental in advancing the automatic detection of intertextuality in Latin and Greek texts. Their open-source tools have seen numerous enhancements and refinements over the years.[2]

Existing research has explored matching words or stems (Coffee et al., 2012) as well as methods that focus on semantics (Scheirer et al., 2014). Additionally, techniques that combine both lexical and semantic elements have been examined, where semantic understanding is established through word embeddings (Manjavacas et al., 2019) or via the (Ancient Greek) WordNet (Bizzoni et al., 2014). While the majority of preceding studies have concentrated on detecting text reuse in the Bible and various religious texts, Burns et al. (2021) focus on Classical Latin literature.

However, to our knowledge, no efforts have been undertaken to automatically detect intertextual similarities across languages, specifically between Greek, Latin, and English texts. This lack is likely due to the inherent complications of inducing cross-language mappings, a difficulty that arises both with surface form-based strategies and with techniques utilizing word embeddings. Notwithstanding, this gap is of significant importance, as it overlooks the frequent appearance of such allusions, especially from Latin to Greek literature.

Language Models for Classical Philology. Bamman and Burns (2020) and Mercelis and Keersmaekers (2022) introduced Latin BERT and ELECTRA models, respectively. For Ancient Greek, Singh et al. (2021) and Yamshchikov et al. (2022) provided BERT models, initialized from Modern Greek BERT and subsequently trained on Ancient Greek data. Similarly, the UGARIT project has successfully explored the usage of the XLM-R model (Conneau et al., 2020) for Ancient Greek and Latin texts (Yousef et al., 2022a,b), even

[2] https://tesserae.caset.buffalo.edu/blog/about-tesserae/.

though XLM-R has not been pre-trained on Ancient Greek texts. Recently, Riemenschneider and Frank (2023) have complemented the encoder-only landscape with encoder-decoder models and developed trilingual models using Ancient Greek, Latin, and English texts. Moreover, Kostkan et al. (2023) and Burns (2023) have developed odyCy and latinCy, respectively, as dedicated spaCy pipelines[3] for Ancient Greek and Latin.

SBERT Embeddings. Reimers and Gurevych (2019) have shown that vanilla BERT embeddings are not suitable for creating sentence embeddings, and instead proposed the S(entence)-BERT models, which are based on Siamese and triplet network structures. Building on their work, Reimers and Gurevych (2020) introduced a method to learn multilingual sentence embeddings via multilingual knowledge distillation. This method realizes knowledge transfer from a monolingual teacher model to a student model, by training the student model to align the original sentence and its translation to the same location in the embedding space.

3 Methodology

We closely follow Reimers and Gurevych's (2020) multilingual knowledge distillation recipe. Their method requires a monolingual teacher model M and parallel sentences in the given source language and the target language(s) $((s_1, t_1), ..., (s_n, t_n))$.

The teacher trains a student model $\hat{M}$ such that $\hat{M}(s_i) \approx M(s_i)$ and $\hat{M}(t_i) \approx M(s_i)$. For a given mini-batch $\mathcal{B}$, the mean-squared loss is minimized:

$$\frac{1}{|\mathcal{B}|} \sum_{j \in \mathcal{B}} \left[(M(s_j) - \hat{M}(s_j))^2 + (M(s_j) - \hat{M}(t_j))^2 \right].$$

In other words, the student model is trained to map a given sentence to the same vector across languages, i.e., the translation of a given sentence should be mapped to the same vector as the source sentence. Notably, this method is not restricted to a bilingual setup. Instead, the student can be trained to map sentence vectors stemming from multiple languages to the same vector, namely the one provided by the teacher model.

In our work, the teacher and student SBERT models to be used for cross-lingual knowledge transfer will be initialized from strong transformer language models for the respective languages. For the English teacher model, we build on the MPNet model of Song et al. (2020), an encoder-only model that has been pre-trained using a combination of masked language modeling and permuted language modeling. Specifically, we use different sentence transformer variants induced from MPNet, as provided by the SBERT library (Reimers and Gurevych, 2019). For the student model, we experiment with initializing it from different multilingual models: XLM-R (Conneau et al., 2020), a multilingual model based on RoBERTa that covers 100 languages, including Modern Greek and Latin, in contrast to PhilBERTa (Riemenschneider and Frank, 2023), a recent trilingual model that has been pre-trained on Ancient Greek, Latin, and English texts.

More detail about our models and the specific experimental setup is provided in Section 5.

4 Parallel Data

As outlined in Section 3, the knowledge distillation method of Reimers and Gurevych (2020) crucially depends on the availability of parallel sentences between the relevant source and target languages – here, the source language English for the teacher model, and English, Ancient Greek, and Latin for our student model.

We collect this data from various sources: from the Perseus Digital Library,[4] from parallel Bible data,[5] parallel English-to-Greek sentences from the OPUS corpus (Tiedemann, 2012), and an extensive collection of parallel English and Latin sentences available on the Huggingface Hub.[6] We refer to the latter dataset as "Rosenthal", named after its associated account.

The Perseus project features a large collection of Ancient Greek and Latin texts, many of them with corresponding translations. However, the alignment of the provided data is not always fine-grained enough for our purpose. Therefore, we align individual lines with their corresponding translation, and discard lines that we cannot align successfully.

To generate additional parallel data for enhanced knowledge transfer, we experiment with translating the English portions of the Rosenthal dataset,

[3] https://spacy.io/.

[4] https://github.com/PerseusDL/canonical-greekLit and https://github.com/PerseusDL/canonical-latinLit.

[5] https://github.com/npedrazzini/parallelbibles/tree/main.

[6] https://huggingface.co/datasets/grosenthal/latin_english_parallel.

	English	Greek	Latin
Perseus	3 743K	2 120K	384K
Bible	897K	128K	520K
Opus	5K	4K	—
Rosenthal	3 428K	2 370K†	2 095K

Table 1: Dataset statistics (in number of words) of available parallel sentences across languages. The Greek Rosenthal data marked with a dagger (†) has been translated using PhilTa$_{\text{En}\to\text{Grc}}$.[7]

which consists solely of English and Latin parallel data, into Ancient Greek. This required first fine-tuning the multilingual PhilTa model[7] on the Perseus data to enable translation from English to Ancient Greek. Subsequently, we used the trained PhilTa$_{\text{En}\to\text{Grc}}$ model to translate the Rosenthal dataset into Ancient Greek, thereby expanding it to a trilingual parallel dataset.

Table 1 provides the data statistics. Since parts of the corpora overlap, we deduplicate the data.

5 Experiments

Our first aim is to compare different model configurations. We test the following configurations:

- **Teacher Model.** We use the `all-mpnet-base-v2`[8] and the `multi-qa-mpnet-base-dot-v1`[9] model from the SBERT library (Reimers and Gurevych, 2019) as teacher models. While the former is fine-tuned on a variety of tasks, the latter is optimized for semantic search.

- **Student Model.** We compare the performance of XLM-R (Conneau et al., 2020) to that of PhilBerta (Riemenschneider and Frank, 2023) when used as student models. XLM-R serves as a well-established multilingual baseline.

- **Data Augmentation.** We evaluate whether the automatic English-to-Greek translations produced by PhilTa$_{\text{En}\to\text{Grc}}$ to extend the Rosenthal dataset improve task performance.

In order to transparently evaluate our models, we first measure their ability to correctly detect translations of a sentence. For each parallel dataset, we hold out 1 000 sentences as test sets. Given a query, i.e., the embedding of a specific sentence in the source language, we compute the cosine similarity to the embeddings of all 1 000 sentences in the target language.

Following Reimers and Gurevych (2020), we measure the success of our models by determining *translation accuracy*: we count a translation to be correctly identified if the model computes the highest cosine similarity between the query and its correct translation, and vice versa. This evaluates the student model's ability to align a source language sentence with an equivalent target language sentence.

However, our primary interest is whether the model can effectively link Ancient Greek and Latin texts. Regrettably, the volume of parallel data available in Ancient Greek and Latin is severely constrained. Consequently, we utilize Bible data, which is accessible in Ancient Greek, Latin, and English. Again, we examine the model's performance on 1 000 test sentences, given in Ancient Greek or Latin. We ensure that the model has not encountered any of these sentences in its training data, either in English or Latin, or in Ancient Greek. In addition, we use the PhilTa$_{\text{En}\to\text{Grc}}$-generated Ancient Greek test set translations of the Rosenthal corpus and compare them to their Latin originals.

We are aware that the task of identifying intertextual allusions poses a much greater challenge than merely recognizing translations, as allusions typically exhibit more subtlety and may extend beyond sentence or verse boundaries. However, we consider this evaluation a transparent method for comparing the effectiveness of different model configurations and an approximate measure to evaluate the potential success of our models in identifying intertextual allusions across languages.

Experiment Details. We train all models with the exact same configurations. We fine-tune all models for 30 epochs, using a batch size of 32, the AdamW optimizer with a learning rate of 2e−5, and 10 000 warmup-steps. The best-performing model is selected based on the translation accuracy derived from a total of 2 000 held-out validation examples, comprised of 1 000 English-Greek and 1 000 English-Latin sentence pairs.

[7] PhilTa (Riemenschneider and Frank, 2023) is a trilingual encoder-decoder model based on T5 (Raffel et al., 2020) that was pre-trained on Ancient Greek, Latin, and English data.

[8] `https://huggingface.co/sentence-transformers/all-mpnet-base-v2`.

[9] `https://huggingface.co/sentence-transformers/multi-qa-mpnet-base-dot-v1`.

Teacher	Student	PHILTA-translations	Bible En→La	Bible La→En	Perseus En→La	Perseus La→En	Rosenthal En→La	Rosenthal La→En
`all-mpnet-base-v2`	XLM-R	✗	0.10	0.10	0.30	0.60	0.50	0.60
`all-mpnet-base-v2`	PHILBERTA	✗	96.10	95.60	90.10	88.40	95.90	95.20
`multi-qa-mpnet`	PHILBERTA	✗	**96.90**	**96.00**	91.60	**91.30**	**97.90**	**96.90**
`multi-qa-mpnet`	PHILBERTA	✓	96.40	95.90	**91.90**	90.90	97.80	96.60

Table 2: Translation accuracy for various *English-Latin* test sets. Utilizing XLM-R as a student model leads to catastrophic results. It is crucial to substitute PHILBERTA as the student model for successful model training. Switching to the semantically-oriented `multi-qa-mpnet` from the broader `all-mpnet-base-v2` provides further enhancements.

Teacher	Student	PHILTA-translations	Bible En→Grc	Bible Grc→En	Perseus En→Grc	Perseus Grc→En	Rosenthal En→Grc†	Rosenthal Grc†→En
`all-mpnet-base-v2`	XLM-R	✗	0.20	0.20	0.30	0.10	0.30	0.10
`all-mpnet-base-v2`	PHILBERTA	✗	96.50	96.50	89.50	87.40	93.39	92.49
`multi-qa-mpnet`	PHILBERTA	✗	97.80	97.70	89.80	88.80	92.29	86.99
`multi-qa-mpnet`	PHILBERTA	✓	**98.30**	**98.00**	**91.10**	**90.50**	**96.80**	**94.29**

Table 3: Translation accuracy for various *English-Greek* test sets. The Greek Rosenthal data has been translated by PHILTA. We see the same trends as in Table 2. The enrichment of the training corpus with additional PHILTA-translated content notably increases the performance for Ancient Greek.

Teacher	Student	PHILTA-translations	Bible La→Grc	Bible Grc→La	Rosenthal La→Grc†	Rosenthal Grc†→La
`all-mpnet-base-v2`	XLM-R	✗	0.10	0.10	0.20	0.20
`all-mpnet-base-v2`	PHILBERTA	✗	96.10	95.60	83.97	83.67
`multi-qa-mpnet`	PHILBERTA	✗	96.50	96.69	84.97	82.57
`multi-qa-mpnet`	PHILBERTA	✓	**96.70**	**96.90**	**92.08**	**91.68**

Table 4: Translation accuracy for various *Latin-Greek* test sets. The Greek Rosenthal data has been translated by PHILTA. We see similar trends as described in Tables 2 and 3.

6 Results

We present our results for the different configurations in Tables 2 to 4. Specifically, we evaluate: i) the performance of different *teacher models* (the more general `all-mpnet-base-v2` SBERT model in comparison to the `multi-qa-mpnet` SBERT fine-tuned for semantic search), ii) different *student models* (XLM-R versus the PHILBERTA model), and iii) *augmenting the parallel data* for training SPHILBERTA using $\text{PHILTA}_{\text{En}\rightarrow\text{Grc}}$-translated texts.

Employing XLM-R as the student model leads to catastrophic performance. Specifically, the model never surpasses the 1% mark in test set performance. We observed this trend consistently, regardless of the model configuration or the random seed employed. This outcome is, to some degree, to be expected, as XLM-R is not pre-trained on Ancient Greek data. Still, it is surprising that XLM-R performs so badly also on Latin data, as its pre-training corpus contained a Latin portion. Moreover, the UGARIT project (Yousef et al., 2022a,b) has successfully adapted XLM-R to Ancient Greek. We hypothesize that the effectiveness of a broadly multilingual but unspecialized model may be task-dependent, and continuing self-supervised pre-training on Ancient Greek texts may be required for XLM-R to adapt adequately. These findings highlight the importance of initializing the student model with a model that is proficient in the target languages.

Initializing the student model with PHILBERTA yields strong performance, often surpassing 95% translation accuracy. Generally, employing `multi-qa-mpnet` as a teacher model contributes to a slight performance improvement over `all-mpnet-base-v2`. Yet, when testing the model on the Ancient Greek Rosenthal corpus, using the `multi-qa-mpnet` teacher model results in a performance decline. Importantly, the Greek part of this dataset has been translated by $\text{PHILTA}_{\text{En}\rightarrow\text{Grc}}$, which could possibly have affected the quality of the dataset. Indeed, while we see this negative trend when *testing* on the generated data, the inclusion of the PHILTA-generated Ancient Greek Rosenthal corpus as additional *training data* leads to a notable enhancement for the Greek datasets, while the performance for Latin translation retrieval remains largely unaffected.

The results for Latin-to-Greek and Greek-to-Latin translations are shown in Table 4. Our models notably exhibit strong performance across both datasets. Again, utilizing the Greek Rosenthal data considerably improves performance. These results show that SPHILBERTA can be efficiently utilized in a scenario that solely involves Greek and Latin texts, without necessitating the involvement of English texts.

7 Case Study: The *Aeneid* and Homer's *Odyssey*

Examinations of the intertextual allusions in Virgil's *Aeneid* to both the *Iliad* and the *Odyssey* have a long history, dating back to antiquity. Structurally, the *Aeneid's* initial six books mirror the narrative of the *Odyssey*, while the concluding six books correspond more closely to the *Iliad*.

In the second book of the *Aeneid*, the protagonist Aeneas attempts to escape from the ravaged city of Troy with his family. Tragically, his wife, Creusa, is lost amidst the chaos. Creusa's ghost consoles him and bids him goodbye before receding into thin air: "*This speech uttered, while I wept and would have said many a thing, she left me and retreated into thin air. Thrice there was I fain to lay mine arms round her neck; thrice the vision I vainly clasped fled out of my hands, even as the light breezes, or most like to fluttering sleep.*"[10]

These verses mirror closely a scene in the Nekyia of the *Odyssey*, where Odysseus meets his mother Anticleia in the underworld: "*So she spoke, and I pondered in heart, and was fain to clasp the spirit of my dead mother. Thrice I sprang towards her, and my heart bade me clasp her, and thrice she flitted from my arms like a shadow or a dream, and pain grew ever sharper at my heart.*"[11]

To evaluate our model's proficiency in identifying these intertextual allusions, we employ each verse of the *Aeneid* passage (i.e., 5 verses) as a query, which we then compare to the verse embeddings (approx. 11 000 verses) of the complete *Odyssey*. Table 5 shows the three highest results for each verse, according to our best-performing model setup (teacher: `multi-qa-mpnet`; student: PHILBERTA; additional PHILTA-generated Rosenthal data).

We note that these verses do not share a direct one-to-one relationship and they are not translations of each other, the scenario in which our model

[10] Virgil, *Aeneid*, 2.790–794, translated by Mackail (1885).

[11] Homer, *Odyssey*, 11.204–208, translated by Murray (1919).

Query	**Results**
Haec ubi dicta dedit, lacrimantem et multa volentem This speech uttered, while I wept and would have said many a thing,	τῆς δ᾽ ἄρ᾽ ἀκουούσης ῥέε δάκρυα, τήκετο δὲ χρώς· and as she listened her tears flowed and her face melted ὣς φάτο, τῆς δ᾽ εὔνησε γόον, σχέθε δ᾽ ὄσσε γόοιο. So she spoke, and lulled Penelope's laments, and made her eyes to cease from weeping. ὣς φάτο, τῇ δ᾽ ἄρα θυμὸν ἐνὶ στήθεσσιν ὄρινε. So he spoke, and stirred the heart in her breast.
dicere deseruit, tenuisque recessit in auras. [...said], she left me and retreated into thin air.	ἡ μὲν ἄρ᾽ ὣς ἔρξασ᾽ ἀπεβήσετο δῖα θεάων, Now when she had done this the fair goddess departed, ἡ μὲν ἄρ᾽ ὣς εἰποῦσ᾽ ἀπέβη πρὸς δώματα καλά, So saying, she departed to the fair palace. ἡ μὲν ἄρ᾽ ἐς κρήνην κατεβήσετο καλλιρέεθρον [She] had come down to the fair-flowing spring [Artacia],
Ter conatus ibi collo dare bracchia circum: Thrice there was I fain to lay mine arms round her neck;	ὄπτ᾽ ἐν χερσὶν ἑλών, τά ῥά οἱ γέρα πάρθεσαν αὐτῷ. he took in his hands roast meat and set it before them, [...] which they had set before himself as a mess of honor. τρὶς μέν μιν πελέμιξεν ἐρύσσεσθαι μενεαίνων, Thrice he made it quiver in his eagerness to draw it, αὐτίκ᾽ ἔπειτα τρίαιναν ἑλὼν χερσὶ στιβαρῇσιν straightway took his trident in his mighty hands,
ter frustra comprensa manus effugit imago, thrice the vision I vainly clasped fled out of my hands,	τρὶς δέ μοι ἐκ χειρῶν σκιῇ εἴκελον ἢ καὶ ὀνείρῳ and thrice [she flitted] from my arms like a shadow or a dream, τρὶς μὲν ἐφωρμήθην, ἑλέειν τέ με θυμὸς ἀνώγει, Thrice I sprang towards her, and my heart bade me clasp her, χερσὶ δὲ μή τι λίην προκαλίζεο, μή με χολώσῃς, But with thy [hands] do not provoke me overmuch,
par levibus ventis volucrique simillima somno. even as the light breezes, or most like to fluttering sleep.	ἡ δ᾽ ἔθεεν Βορέῃ ἀνέμῳ ἀκραέϊ καλῷ, And she ran before the North Wind, blowing fresh and fair, ὄρσας ἀργαλέων ἀνέμων ἀμέγαρτον ἀυτμήν, when he had roused a furious blast of cruel winds ἐς πνοιὰς ἀνέμων. ἡ δ᾽ ἐξ ὕπνου ἀνόρουσε into the breath of the winds. And [she] started up from sleep

Table 5: Top 3 predictions of our best-performing SPHILBERTA model (teacher: `multi-qa-mpnet`; student: PHILBERTA; additional PHILTA-generated Rosenthal data) when queried over the whole *Odyssey*. We mark corresponding cross-lingual concept pairs with individual colors.

was trained. Even so, we observe that verse 793 ("*thrice the vision I vainly clasped fled out of my hands*") is correctly paired with the best corresponding Greek verse ("*and thrice she flitted from my arms like a shadow or a dream*"). In the majority of cases, our model accurately captures crucial concepts, such as *weeping*, *departing*, *triplicity*, *wind*, and *sleep*, linking them reasonably to different verses. However, our verse-to-verse mapping, which precludes longer texts, results in a lack of a cohesive concept of extended intertextually alluding passages.

Still, our case study demonstrates the proficiency of our models in recognizing sentence structures and translating them to a different language (as in "*this speech uttered*" → "*so she spoke*"), and in identifying common topics or concepts across languages, even locating verses where multiple relevant concepts exist within the same verse ("*thrice*", "*the vision*", "*out of my hands*" → "*thrice*", "*a shadow or a dream*", "*from my arms*").

Despite these remarkable results, our case study also reveals the need for a more sophisticated retrieval mechanism that extends beyond verse boundaries to search for broader patterns. Yet, already in the present state, our SPHILBERTA model can serve as a useful tool for automatic first-pass exploration of potential cross-lingual intertextual allusions, and in this way can support philologists in the search for intertextual allusions.

8 Conclusion

We introduce SPHILBERTA, a multilingual PHILBERTA-derived sentence transformer model, specifically adapted to Classical Philology. Our model represents a pioneering effort in detecting intertextual allusions between Ancient Greek and Latin texts, which is characteristic of many Roman writers who used Greek literature for juxtaposition. SPHILBERTA displays impressive performance across various datasets, confidently identifying direct translations among English, Latin, and Ancient Greek. We have illustrated that SPHILBERTA holds strong potential in revealing intertextual allusions; however, additional research is needed to fully exploit the model's capabilities. Our multilingual SPHILBERTA and the similarity-driven retrieval settings built upon it offer, for the first time, the option to study intertextuality cross-lingually on a broader scale in original Classical Literature.

Acknowledgements

We thank Nina Stahl for the insightful discussions we shared on the Greek and Latin parallel passages.

References

María Jesús Martínez Alfaro. 1996. Intertextuality: Origins and development of the concept. *Atlantis*, 18(1/2):268–285.

David Bamman and Patrick J Burns. 2020. Latin bert: A contextual language model for classical philology. *arXiv preprint arXiv:2009.10053*.

Andreas Bendlin. 2006. Intertextuality. `http://dx.doi.org.ubproxy.ub.uni-heidelberg.de/10.1163/1574-9347_bnp_e525570`. Accessed: 28 June 2023.

Yuri Bizzoni, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini, and Gregory Crane. 2014. The making of Ancient Greek WordNet. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1140–1147, Reykjavik, Iceland. European Language Resources Association (ELRA).

Patrick J Burns. 2023. Latincy: Synthetic trained pipelines for latin nlp. *arXiv preprint arXiv:2305.04365*.

Patrick J. Burns, James A. Brofos, Kyle Li, Pramit Chaudhuri, and Joseph P. Dexter. 2021. Profiling of intertextuality in Latin literature using word embeddings. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4900–4907, Online. Association for Computational Linguistics.

Neil Coffee, Jean-Pierre Koenig, Shakthi Poornima, Christopher W Forstall, Roelant Ossewaarde, and Sarah L Jacobson. 2012. The tesserae project: intertextual analysis of latin poetry. *Literary and linguistic computing*, 28(2):221–228.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Christopher Forstall, Neil Coffee, Thomas Buck, Katherine Roache, and Sarah Jacobson. 2014. Modeling the scholars: Detecting intertextuality through enhanced word-level n-gram matching. *Digital Scholarship in the Humanities*, 30(4):503–515.

G. O. Hutchinson. 2013. *Greek to Latin: Frameworks and Contexts for Intertextuality*. Oxford University Press.

Jan Kostkan, Márton Kardos, Jacob Palle Bliddal Mortensen, and Kristoffer Laigaard Nielbo. 2023. OdyCy – a general-purpose NLP pipeline for Ancient Greek. In *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 128–134, Dubrovnik, Croatia. Association for Computational Linguistics.

Julia Kristeva. 1986. Word, dialogue, and novel. In *The Kristeva reader*, pages 34–61. Columbia University Press.

John William Mackail. 1885. *The Aeneid of Virgil*, volume 36. Macmillan.

Enrique Manjavacas, Brian Long, and Mike Kestemont. 2019. On the feasibility of automated detection of allusive text reuse. In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 104–114, Minneapolis, USA. Association for Computational Linguistics.

Wouter Mercelis and Alek Keersmaekers. 2022. An ELECTRA model for Latin token tagging tasks. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, pages 189–192, Marseille, France. European Language Resources Association.

A. T. Murray. 1919. *Homer. The Odyssey with an English Translation*. Harvard University Press, London.

Mary Orr. 2003. Intertextuality: Debates and contexts. *Polity Press*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4512–4525, Online. Association for Computational Linguistics.

Frederick Riemenschneider and Anette Frank. 2023. Exploring large language models for classical philology. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15181–15199, Toronto, Canada. Association for Computational Linguistics.

Walter Scheirer, Christopher Forstall, and Neil Coffee. 2014. The sense of a connection: Automatic tracing of intertextuality by meaning. *Digital Scholarship in the Humanities*, 31(1):204–217.

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.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: Masked and permuted pre-training for language understanding. *arXiv preprint arXiv:2004.09297*.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2214–2218, Istanbul, Turkey. European Language Resources Association (ELRA).

Ivan Yamshchikov, Alexey Tikhonov, Yorgos Pantis, Charlotte Schubert, and Jürgen Jost. 2022. BERT in plutarch's shadows. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6071–6080, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Tariq Yousef, Chiara Palladino, Farnoosh Shamsian, Anise d'Orange Ferreira, and Michel Ferreira dos Reis. 2022a. An automatic model and gold standard for translation alignment of Ancient Greek. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5894–5905, Marseille, France. European Language Resources Association.

Tariq Yousef, Chiara Palladino, David J. Wright, and Monica Berti. 2022b. Automatic translation alignment for Ancient Greek and Latin. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, pages 101–107, Marseille, France. European Language Resources Association.