

Using Parallel Corpora to Evaluate Translations of Ancient Greek Literary Texts

An Application of Text Alignment for Digital Philology Research

Chiara Palladino¹ 
Farnoosh Shamsian² 
Tariq Yousef² 

1. Classics, Furman University , Greenville, USA.
2. Leipzig University , Leipzig, Germany.

Citation

Chiara Palladino, Farnoosh Shamsian, and Tariq Yousef (2022). "Using Parallel Corpora to Evaluate Translations of Ancient Greek Literary Texts. An application of Text Alignment for Digital Philology Research". In: *Journal of Computational Literary Studies* 1 (1). 10.48694/jcls.100

Date published 2022-12-13

Date accepted 2022-04-04

Date received 2021-12-21

Keywords

translation alignment, translation analysis, Philology, Critical Translation Studies, NLP, annotation

License

CC BY 4.0 

Reviewers

Marianne Reboul 
Matteo Romanello 

Note

This paper has passed through the conference track of JCLS. In addition to being peer reviewed, it was presented and discussed at the 1st Annual Conference of Computational Literary Studies 2022 at the Technische Universität Darmstadt.

Abstract.

This paper presents a workflow to systematically compare translations of Ancient Greek into English and Persian through the analysis of parallel corpora aligned manually at word level. We extracted the translation pairs, measured word intersections, alignment types, and part of speech matches, in order to investigate quantitative indicators of closeness to the original and similarity across translations. The corpus includes passages from the *Iliad* and the *Hippolytus* by Euripides. In addition to direct translations, we have included some indirect translations of the *Iliad* in Persian, where French was used as a mediating language.

1. Introduction

In this article, we propose an application of translation alignment for the study of translations of Ancient Greek texts in English and Persian. We introduce the general principles of translation alignment and its challenges in the domain of historical languages, and examine how the alignment of parallel texts at word level can support a comparative analysis of translations and the identification of certain translation phenomena.

The operation of aligning texts in different languages is called Translation Alignment: It is one of the most important tasks of Natural Language Processing, and it essentially consists in establishing correspondences between two or more texts. Such correspondences can be at various levels: document, page, paragraph, sentence, or word/token (Kay and Röscheisen 1993). Texts aligned at any level of granularity are defined parallel texts or parallel corpora (Véronis 2000). Various methods have been developed to perform translation alignment automatically: from the more traditional statistical models like Giza++ (Och and Ney 2003), which utilize large amounts of annotated training datasets, to more recent and innovative contextualized language models (Dou and Neubig 2021; Jalili Sabet et al. 2020; Yousef et al. 2022c).

High-quality parallel texts aligned at word level, whether manually or automatically, are an essential resource that is used in several fields: Primarily, they provide training data for statistical methods, or gold standards against which contextual models can



Figure 1: The home page of Ugarit.

be tested. However, they are also used in many other domains, including text mining, pedagogy, and text reuse detection (Dagan et al. 1999; Graça et al. 2008; Palladino et al. 2021). Because of this intrinsic importance, there are several tools that are designed to facilitate the user-based creation of parallel texts at word and sentence level: Some tools offer an annotation interface without visualization of the completed alignments (Caseli et al. 2002; Grimes et al. 2010; Melamed 1998), while others empower various kinds of visualizations and queries on annotated texts (Almas and Beaulieu 2013; Barreiro et al. 2016; Germann 2008), and may provide support for additional annotations, such as part-of-speech and syntactic dependencies (Gilmanov et al. 2014).

1.1 Introducing Ugarit: A Tool for Translation Alignment of Low-Resourced Languages

Automatic alignment models tend to perform poorly with low-resourced languages, because they often require large quantities of already aligned texts as training data, or millions of words in digitized corpora. This is an issue especially for historical languages, which do not have the same NLP infrastructure as modern ones, and have also a limited amount of fully digitized texts. For this reason, the creation of high-quality aligned corpora is of paramount importance.

The tool used for this study, Ugarit, is a web-based Translation Alignment editor designed with ancient or low-resourced languages in mind (<http://ugarit.ialigner.com/>, see also Figure 1). It is a crowd-sourcing project that enables users to align up to three parallel texts at sentence or word level, specifically focusing on texts less represented in the field of computer-based alignment.

The workflow is simple: The user uploads the desired texts or imports them from the Perseus Digital Library, and clicks on the words to align, which are then stored in the database as translation pairs. A progress bar allows the user to see how much has been aligned. Users can create translation pairs aligning one word to another word (1-1), one word to many words (1-N), many words to one (N-1), and many to many (N-N). By default, the alignments are published on the platform in the “New Alignments”

Blind Owl, Sadegh Hedayat, First Paragraph by Costello
Farnoosh Shamsian /

Created on 2022-04-06 16:56:59 Modified on 2022-04-07 10:48:35 Translated by D.P. Costello Aligned by Farnoosh Shamsian

Transliterate فارسی English

در زندگی زخماهایی هست که مثل زخمی در انزوا روح را آهسته در انزوا میخورد و میترشد . این دردها را نمی‌شود به کسی (شماره ۱۰۵) . چون عموماً عادت دارند که این دردهای باورنکردنی را جزو اتفاقات و پیشامدهای ناظر و عجیب بشمارند و اگر کسی بگوید یا بنویسد ، مردم بر سبیل عقاید جاری و عقاید خدیشان سعی میکنند آن را با ایخند شکاک و تمسخرآمیز تلقی کنند . زیرا پشتر هنوز چاره و دوائی برایش پیدا نکرده و تنها داروی آن فراموشی توسط شراب و خواب مصنوعی به وسیله اقیون و مواد مخدره است- ولی افسوس که تاثیر این کینه دارو ها موقت است و به جای تسکین پس از مدتی بر شدت درد میافزاید . آیا روزی به اسرار این اتفاقات ماوراء طبیعی ، این انعکاس سایه روح که در حالت اغما و بیزخ بین خواب و بیداری جلوه می کند کسی پی خواهد برد؟

There are sores which slowly erode the mind in solitude like a kind of canker . It is impossible to convey a just idea of the agony which this disease can inflict . In general , people are apt to relegate such inconceivable sufferings to the category of the incredible . Any mention of them in conversation or in writing is considered in the light of current beliefs , the individual's personal beliefs in particular , and tends to provoke a smile of incredulity and derision . The reason for this incomprehension is that mankind has not yet discovered a cure for this disease . Relief from it is to be found only in the oblivion brought about by wine and in the artificial sleep induced by opium and similar narcotics . Alas , the effects of such medicines are only temporary . After a certain point , instead of alleviating the pain , they only intensify it . Will anyone ever penetrate the secret of this disease which transcends ordinary experience , this reverberation of the shadow of the mind , which manifests itself in a state of coma like that between death and resurrection , when one is neither asleep nor awake ?

(44) 30% FAR (105) 70% FAR - ENG (126) 61% FAR - ENG (79) 39% ENG

Figure 2: Public view of an alignment, showing the transliteration feature.

panel, although users may opt out by simply selecting a different visibility option. The translation pairs can be further examined using the Alignment Statistics chart, which counts the frequency of the types of pairs created, or by downloading the whole dataset in XML or tabular format. It is also possible to analytically inspect published alignments by hovering with the mouse on each token: Aligned words and phrases are highlighted in both texts. An additional service of transliteration is also provided for most of the languages in the database (see Figure 2).

Ugarit stores translation pairs as a graph database, which can be used for dynamic lexica induction. Further, Ugarit allows users to inspect how other people aligned a specific word using the translation pairs search functions, which provide a contextualized visualization of an aligned pair (Yousef et al. 2022b).

Because of these powerful supporting features, Ugarit has been variously used for research, machine translation development, and language learning (Crane et al. 2019, Yousef et al. 2022b, Shukhoskvili 2017, Yousef et al. 2022a, Yousef et al. 2022c). The user pool currently counts 581 users, and more than 40 different languages, including Ancient Greek, Persian, Latin, Egyptian, Coptic, Arabic, Georgian, and Akkadian, and more than 250,000 texts have been aligned by scholars, teachers, and students.

2. Methodology

The use of text alignment for the analytical study of translations is a relatively new idea. In the field of modern languages, a line of inquiry called Corpus-Based Translation Studies (CTS) uses parallel corpora, coupled with information like syntax dependencies and sentiment analysis, to achieve a better understanding of translational dynamics or textual traditions (Baker 1993; Laviosa 2008).

For historical languages, relatively little has been done in this area, although the study of translations of ancient literary works is of unquestionable importance (Bettini 2012; Nergaard 1993). Bizzoni et al. (2017) used an automatic aligner based on the Needleman-Wunsch algorithm to create a large parallel corpus of French translations of the *Odyssey* and identify diachronic trends across them. Their method was based on the extraction of specific aligned passages anchored to proper nouns, samples of which were

then evaluated manually and used to identify trends in translation practices across a chronological span (16th-17th century).

Our approach goes partly in the opposite direction, and it attempts to apply an analytical method of digital close reading through annotation. Ugarit has variously demonstrated the potential of translation alignment in analytical tasks on texts and languages, based on the reflective evaluation of correspondences between words (Palladino 2020). Therefore, it can also be used for the analytical study of translations of ancient texts, supporting fine-grained research questions that require a certain control of the data. For example, it can enable researchers to establish phrasal correspondences based on the peculiarities of a text or language. Moreover, it empowers the study of aligned corpora in languages that are currently unsupported by NLP architectures, such as Persian, as demonstrated in our study.

In addition, we also attempt to find quantitative indicators that may help in the cumulative evaluation of translations after the completion of the annotation stage. This double approach ensures that, on the one hand, the researcher has full control on the data creation stage, and on the other hand, it introduces external criteria to validate or problematize their hypothesis.

We selected two texts from the Ancient Greek tradition, Euripides' *Hippolytus* and the *Iliad*, and aligned them against translations in English and Persian, respectively. The rationale in the selection of the texts was largely due to availability in digital format, although both texts retain much interest for their humanistic value. The *Hippolytus*, a tragedy written by Euripides in 428 BCE based on a previous version now lost, is well-known for its problematic character, dealing with subjects like incest and misogyny. For this study, we selected two famous passages, the prologue of Aphrodite (vv. 1-20) and the anti-women tirade of Hippolytus himself (vv. 616-638), which exhibit similar features in semantics and vocabulary.

Choosing the *Iliad* perhaps does not require justification, but it was also due to external factors. For Persian, the scarcity of direct translations from Ancient Greek is the main challenge, as most translations are indirect and derived from mediating translation(s). Although most Ancient Greek texts have not only one, but multiple indirect translations in Persian, we wanted to include at least one direct translation, which limited the range of choices. The *Iliad* is one of the very few texts that, in addition to two indirect translations, also has a direct translation in Persian. Using translation alignment for a comparison between indirect and direct translations gives us practical information for evaluating accuracy and reliability. Considering that indirect translations are the main method for transmission of the Ancient Greek texts to Persian, the question of their accuracy is of great significance. In addition to the direct translation of the *Iliad*, we also have included two indirect translations, both based on French translations. Moreover, the three translations of the *Iliad* come from different backgrounds and therefore show sufficient variation for testing our methodology.

The alignment was conducted on Ugarit by one annotator for English and one for Persian. Each annotator completed the alignments following a previously established set of guidelines that had been used to train a contextual model (Yousef et al. 2022a). These

guidelines were expanded to increase tolerance to phrasal and idiomatic constructs, while overall retaining the basic principles.¹

We then collected the resulting Translation Pairs (TPs): While the interface provides a local option to download this information, we extracted it directly from the database to optimize data cleaning. Then, we measured several variables: 1) lack of alignment; 2) word intersection across translations; 3) link types; 4) for English, we were also able to calculate matches in part-of-speech.

This study is conducted on a small dataset, and it cannot be considered an extensive analysis on the textual tradition of *Hippolytus* or the *Iliad*. However, it serves as a demonstration of the methodology and as an assessment of the quantitative indicators that can be applied to measure the performance of a translation, as a result of the operation of translation alignment. In the conclusions, we evaluate the advantages and obstacles to this approach, and discuss its potential expansion to larger datasets.

3. Alignment of Euripides, *Hippolytus*

The user compared four competing translations of the Greek tragedy *Hippolytus*:

- D. Kovacs, Euripides. *Children of Heracles. Hippolytus. Andromache. Ecuba*. Loeb Classical Library, Cambridge: Harvard University Press, 1995. This is one of the most recent scholarly editions, featuring the original facing the translation, and praised for its grammatical accuracy (Gibert 2022).
- G. Theodoridis, Euripides, Volume Three. *Medea, Herakleidae, Herakles, Hippolytus.*, 2010. The only translation conceived as a script for a theatrical performance, rather than for reading.²
- I. Johnston, Euripides, *Hippolytus*, 2016. A translation written specifically for the general public, including teachers and students and the only one in poetry.³
- E.P. Coleridge, *The Plays of Euripides. Translated into English Prose from the Text of Paley*, G. Bell and Sons: London 1910. Commissioned as a prose translation by the publisher, it was delivered by the translator with the intent of being “an accurate rendering of the Greek text with some elegance of expression” (preface, p. 11).⁴

3.1 Analysis of Non-Aligned Words

The visualization of the alignments in Ugarit provides a nice overview of the most visible characteristics of each translation, alongside a quick glance on the percentage of aligned and not aligned tokens.

1. The result of this change was that the number of phrase-to-phrase alignments, which were considered less useful for computational implementations, increased in the context of this study.

2. Made available for use on <https://bacchicstage.wordpress.com/euripides/hippolytus/> (accessed Nov. 20, 2022).

3. Made available for use on <https://johnstoniatexts.x10host.com/euripides/hippolytushtml.html> (accessed Nov. 20, 2022).

4. Available on WikiSource: [https://en.wikisource.org/wiki/The_Plays_of_Euripides_\(Coleridge\)](https://en.wikisource.org/wiki/The_Plays_of_Euripides_(Coleridge)) (accessed Nov. 20, 2022).

Translation	NA in Greek	NA in English	L1/L2 Rate
Kovacs	8.19%	6.9%	0.5714
Theodoridis	15.95%	31.24%	0.3939
Johnston	11.64%	13.32%	0.5066
Coleridge	7.33%	4.86%	0.5934

Table 1: Percentage of non-aligned words (NA) in the translation and in the original in both sections of *Hippolytus*, and length rate between original and translation.



Figure 3: A chart representing the proportions and exact number of non-aligned words in all four translations.

Visualized alignments also make it easier to identify overarching tendencies in non-aligned words. In our estimates, we have excluded punctuation to provide more informational numbers. We also provided the length rate calculated as the relation between the number of words in the original and the total number of words in the translation (see [Table 1](#) and [Figure 3](#)).

Lack of alignment in Ugarit indicates that the user, following the guidelines, did not find an acceptable correspondence for a word or phrase. As a working assumption based on previous experience, we hypothesized that an equivalently high number of non-aligned words in both languages would indicate some lack of overlap between original and translation. This, however, occurs rarely in scholarly translations. On the other hand, there may be a strong imbalance in the number of non-aligned words in either language: if a large number of words in Ancient Greek were not aligned, for example, this may indicate a tendency to omission. If a larger number of words in English were not aligned, this may be an indicator of expansion or paraphrasing.

Globally, the percentage of non-aligned words in Greek was between 7 and 15%. In English, the values oscillated much more: from 4.8 in Coleridge, up to 31% in Theodoridis. The length rate between the two texts was also approximately above 50% in three out of four translations, meaning that for every word in the original, a translation would use

1.5 words or more. The fourth translation, Theodoridis, had a much lower length rate, indicating that this translation would be almost three times longer than the original.

We inspected the dataset, to verify what kinds of words were omitted, and enriched this analysis with part-of-speech tagging (on which see below). As it is to be expected, most translators omit functional words in Ancient Greek, such as particles, adpositions, conjunctions, and adverbs. So, words such as μέν, τε, καί, δέ, δή, ἄν, γάρ, which do not have a proper English translation, are often omitted or rendered in other ways.

On the other hand, the types of words that were most frequently added in English also covered specific grammatical functions: Pronouns, for example, are a necessary feature in English but are typically omitted in Greek. Adpositions, adverbs and conjunctions are also frequently added.

There are certain features, however, that stand out. Theodoridis is definitely the author who omits the most parts of the original, and, more notably, who adds English words at a staggering ratio. This translation tends to omit important parts of speech (e.g. nouns and concepts: βάρος, γυνή, κακόν, καλόν, γένος, ἀλήθεια), and to add a very large number of English words and concepts, including nouns, verbs, adjectives, and even proper nouns. This suggests that, while a substantial part of the original was left out, there is a strong counter-tendency to expand on the Greek text.

The two translators that retain a very balanced rate of non-aligned words are Kovacs and Coleridge. Kovacs omits little more than particles and adverbs, and makes significant additions in English only in a few cases. Perhaps Coleridge shows the most unexpected result, being the translation with the lowest and most consistent non-alignment rates. In only two cases the omission of the Greek is particularly relevant, both in vv. 1-20: The expression “κοῦκ ἀνώνυμος”, referred to Aphrodite (lit. “not anonymous”, famous), which is paraphrased and incorporated in the rest of the verse “wide o’er man my realm extends, and proud the name”, and the verb “ψάθει” (lit. “touches”), which is replaced in context with “will (have) none of it”. In the remaining occurrences, most of the words omitted are stopwords or redundancies (e.g. “μύθων”, “of words”, is omitted from the expression “the truth of this (scil. of these words)”).

Somehow more surprisingly, the result is much less positive for the other modern translator, Johnston, who is only second to Theodoridis in both omissions and additions. Johnston also tends to omit some relevant words in Greek (e.g. γῆ, πολίτης), but especially shows a tendency towards expansion in English, with the addition of significant words and concepts that tend to be explanatory of the Greek (e.g. “the would-be husband”, “wife”, “worthy family”, “Hippolyta”, “women”, “time”, etc.).

3.2 Similarities: Analysis of Intersection Data

We extracted intersection data from all four alignments, then compared intersections across all translations and across any combination of them. We observed that literal intersection between each pair of translations is *always* minimal. Overall, among all translations, intersection was found in 15 words in total after capitalization, but before lemmatization. These were the following: ἡμᾶς – me, δ’ – but, τ’ – and, τὰμά – my, κράτη – power, Ἄρτεμιν – Artemis, κόρην – daughter, ἀδελφήν – sister, ἢ – or, ἐν – in, σίδηρον – iron, τὶ δὴ – why, εἰ – if, γυναικῶν – women, τε καί – and.

Non-Aligned words in Ancient Greek, classified by part-of-speech

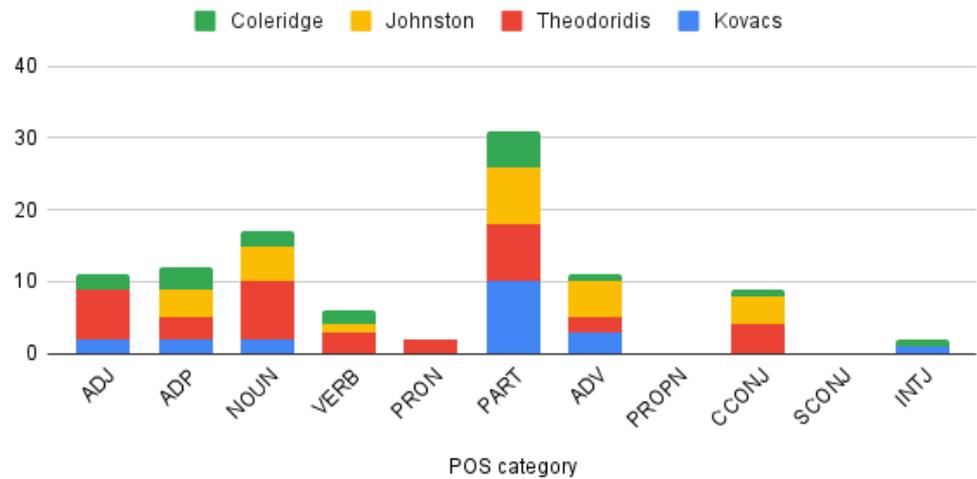


Figure 4: Distribution of non-aligned Greek words in the four translations, classified by POS category.

Non-Aligned words in English translations of Hippolytus, classified by part-of-speech

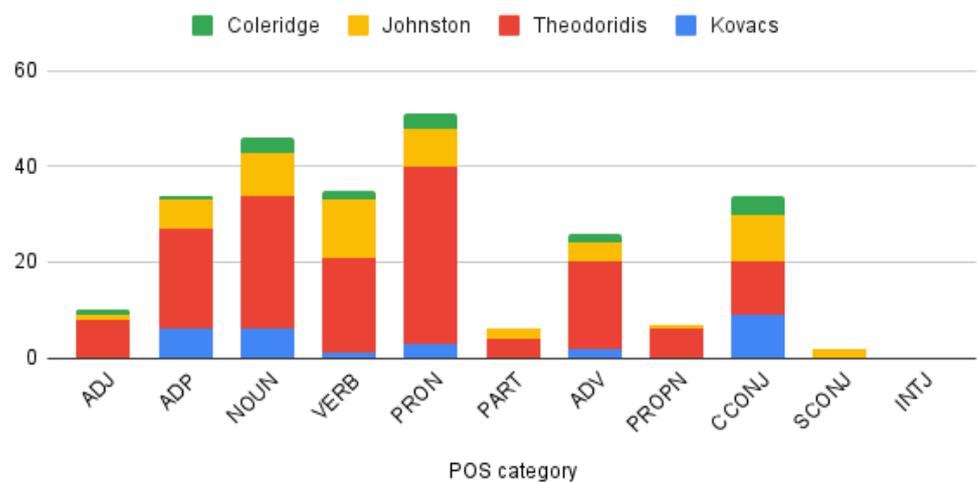


Figure 5: Distribution of non-aligned English words in the four translations, classified by POS category.

Translation pair	Intersection rate
Kovacs – Johnston	18.0%
Kovacs – Coleridge	16.7%
Theodoridis – Johnston	9.2%
Theodoridis – Coleridge	8.9%
Theodoridis – Kovacs	11.1%
Johnston – Coleridge	10.0%
<hr/>	
Total intersection	5.2%
	(12.2% after normalization)

Table 2: Word intersection rates across the four translations of *Hippolytus*. The last line includes the change in percentage after the normalization tasks described.

The intersection is not only minimal, but also relatively insignificant as to the typologies of overlapping words, which include adpositions, particles, and conjunctions. We expanded this list through lemmatization, normalization, and by whitelisting articles preceding proper nouns, which are commonly used in Greek but never in English. With this increased level of tolerance, we included the following words: θεά – goddess, Ζεύς – Zeus, Τροζηνία – Troizen, Πιτθεύς – Pittheus, Ἴππόλυτος – Hippolytus, ὄλβος – wealth, χρυσός – gold, γυνή – woman, γένος – race, λέκτρον – wife, φερνή – dowry (but note ‘dower’ in Coleridge).

Finally, in a number of cases we had a 3/4 overlap and the fourth translation was only minimally different. These include: ἕκαστος – each man, ἐλεύθερος – free, χαλκός – bronze, βροτός – man, χρηστός – good, Ἀμαζών – Amazon (note Kovacs, ‘Amazon woman’), ἐγώ – I, θεός – god.

Additional patterns of overlap can be seen across each pair of translation, as shown in [Table 2](#).

At close inspection, we observed the most frequent overlap in the following semantic categories:

- Proper nouns: Ἄρτεμις – Artemis, Ζεύς – Zeus, Ἀμαζών – Amazon, Θησεύς – Theseus, Πιτθεύς – Pittheus, Ἴππόλυτος – Hippolytus;
- Functional words: prepositions, adverbs, conjunctions, and pronouns;
- Common or very common words that have a standardized translation in a given context: In our case, these pertained to the religious or family sphere: ὄλβος – wealth/family wealth, δῶμα – home/house/estate, θρέπω – to raise, φερνή – dowry, γυνή – woman, βρότειος/βροτεία/βροτός – mortal, ναός – temple/shrine, οὐρανός – heaven, σέβω – to respect (a deity), πατήρ – father, πενθερός – in-law, τιμάω – to receive honors/be revered;
- Technical or rare words that have few established meanings. In our case, this was especially names of metals: σίδηρος – iron, χαλκός – bronze, χρυσός – gold.

Even in these cases, however, some words belonging to these categories did not overlap. The most prominent case is Πόντος, a name that simply stands for “sea”, but conventionally refers to the Black Sea when capitalized. Kovacs and Johnston both opted

for the Latinized name Euxine Sea (Pontus Euxinus), more traditional in scholarship, and Theodoridis preferred the modern version “Black Sea”, while Coleridge simply translated “the sea”.

For the two translations that have the highest intersection, Kovacs and Johnston, there is a high level of literal overlap in all the categories described above: We observed a total of 22 functional words (prepositions, conjunctions, and adverbs), 6 proper names, and 11 family or religious names. There were also many cases of conscious stylistic choices: Both authors translated Κόπρις as “Aphrodite”, Πόντος as “the Euxine Sea”, and the genitive Ἀτλαντικῶν as “(the Pillars) of Atlas”. Certain common words were translated in the same way: In particular, the gen. οὐρανοῦ was translated by both authors as a locative (“in heaven”) and the word φυτόν with the neutral “creature”, as opposed to Theodoridis and Coleridge, who chose derogatory words (“beast” and, notably, “weed”).

While Coleridge did not fare as bad as we expected, the overlap is limited to functional words and a few common categories as observed above. On the other hand, Coleridge distinguished himself in some cases: οὐρανός – “heaven’s courts”, as opposed to the prevalent “heaven”; acc. of motion χλωρὰν ὕλην (lit. “through the green forest”) – “through the Greenwood”; λέκτρον (lit. “the marriage bed”) – “Love” (note capitalization); ἥλιος (lit. “sun”) – “the sun-god”; δῶμα (lit. “house”) – “independence”; λῆχος (lit. “marriage bed” or “marriage bond”) – “wife”. These results are not surprising, considering the age of the translation: Coleridge’s wording may be more distinctive than the rest of the group, and consequently show less overlap.

Theodoridis, on the other hand, regularly scores low intersections despite being a very modern translation. This may be explained by the very different audience of this translation, which was written as a script rather than for reading. One exception is the overlap with Kovacs in the notable translation of λέκτρον as “the bed of love”, which is unique in our group. However, Theodoridis makes very distinctive choices in cases where the other three show strong semantic overlap: παρθένος (lit. “virgin”) - “little virgin deity”, Θεσέως (lit. “of Theseus”) - “by the seed of Theseus”, ναίω (lit. “I dwell”) - “live out their lives”, σπείρω - “to sow the seeds”, ἐκπονεῖ (lit. “he works out, finishes off”) - “he begins the little game of cajoling” (sic!).

Some parts of the text are consistently translated in drastically different ways. This was most commonly observed in idiomatic constructs and fixed expressions, which were addressed differently by each scholar. For example, “φρονοῦσιν μέγα”, lit. “they think great things”, was only translated literally by Kovacs (“think proud thoughts”), but it was grammatically altered or paraphrased by the others: “treat with disrespect” (Theodoridis), “stuffed with pride” (Johnston), “vaunt themselves” (Coleridge). More conventional expressions were also translated in different ways: For example, “ἔχει δ’ ἀνάγκην” is an idiomatic construct meaning “there is a necessity”, and it was often expanded to better render the semantic depth of the word ἀνάγκη, which is very important in the Greek tragic vocabulary (Munson 2001): “there is a fatal necessity” (Kovacs), “and then come the inavoidable choices of his constraints” (Theodoridis), “(he) has a fatal choice” (Johnston), “for he is in this dilemma” (Coleridge).

Translation	Matching POS	Partial matching POS	Non-Matching POS
Kovacs	68.3%	11.0%	8.2%
Theodoridis	66.0%	8.2%	5.0%
Johnston	69.3%	10.5%	6.4%
Coleridge	69.3%	11.9%	9.1%

Table 3: Table of overlapping POS across the four translations.

However, we observed the widest and most regular disagreement in one single word: The neutral (τὸ) κακόν, which appears repeatedly in our verses as a signpost for “woman”. Each translator used a noticeable variety of synonyms, including “curse”, “evil”, “bane”, “problem”, “plague”, “trouble”, “unbearable burden”, “mischief”, “worthless”, and “brainless figurine” (sic!), with remarkable variety shown in the space of about forty verses. Other conventional words often repeated in the dataset, such as “woman” (γυνή), “power” (κράτη) or “mortal” (βροτός), were translated with much less creativity.

3.3 POS Data

The comparison of part-of-speech data is insufficient to reveal significant trends. At an earlier stage of the study, we ran a POS tagger on the Greek and on the English to detect matches between them. We used UDPipe (<https://lindat.mff.cuni.cz/services/udpipe/>) trained on the Perseus Ancient Greek corpus, and obtained accurate enough results that required minimal manual revision. However, as it is to be expected, a single POS in Greek almost always corresponds to multiple POS in English, so we added a further stage of data cleaning and queried partial matches, i.e. TPs that included the same POS in both languages. Finally, we also included cases where the discrepancy in POS was not relevant as partial matches, e.g. a particle in Greek that can only be translated as an adverb in English.

Overall, the resulting data (see Table 3) revealed that in the vast majority of cases, at least one equivalent POS to the original ancient Greek is contained in the corresponding translation, with the most prevalent categories being nouns and verbs. Considering that the analysis of non-aligned tokens (above) is complementary to this result, this was not surprising.

As for the non-matching POS, the data is insufficient to come to decisive conclusions, but there were no clear trends, presumably due to the small size of the dataset. The most common non-matching pairs were the following: VERB-NOUN, NOUN-ADP, ADV-NOUN, NOUN-ADP, PRON-NOUN, ADJ-NOUN, NOUN-VERB, PRON-ADV, PART-ADV. The least prevalent matching POS included pronouns, conjunctions, particles, and pronouns, with values between 2% and 5%.

Nothing immediately visible characterized the more fluent translations, such as Theodoridis, or the more literal ones, like Kovacs or Coleridge. The rate of non-matching POS in Theodoridis seems lower than the others; however, that is also due to the fact that many words in this translation are simply not aligned. Therefore, the rate of POS matches is not a reflection of higher accuracy, but simply of the occurrences where some already aligned words have the same grammatical role.

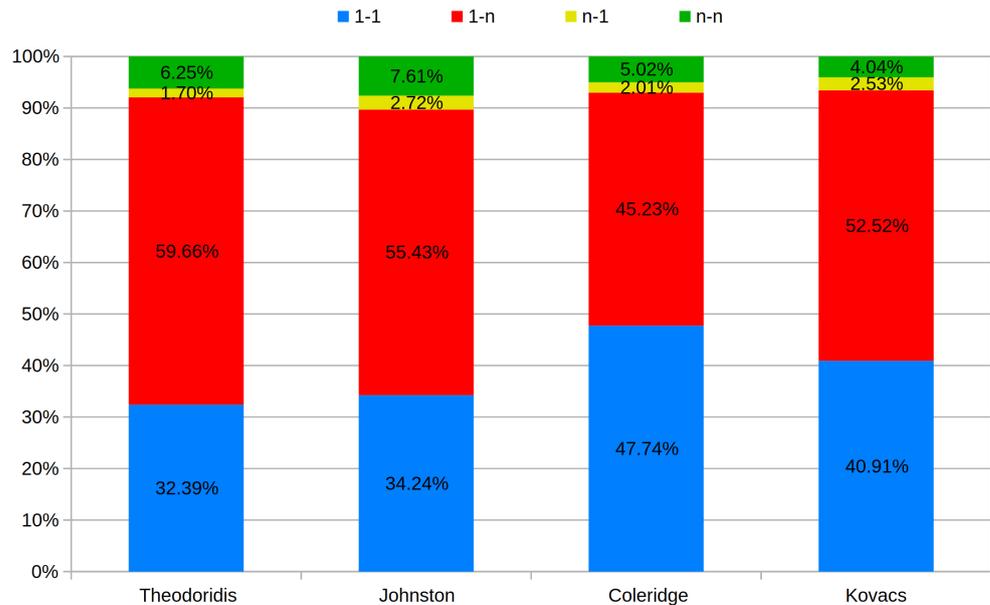


Figure 6: Types of translation pairs across the four translation of *Hippolytus*.

3.4 Types of Translation Pairs

Figure 6 and Figure 7 show the ratios of Translation Pairs (TPs) classified by link types across the four translations: 1-1 (word-to-word), 1-N (word-to-phrase), N-1 (phrase-to-word) and N-N (phrase-to-phrase). In general, these ratios are consistent across the group, with about a third of 1-1 links, regularly higher in Coleridge and Kovacs. The former, however, surpasses the latter by almost 10%: This means that single-word overlap is much more frequent in Coleridge than in the more modern Loeb translation.

1-N links are also regularly between 50% and 60% of the total in all four: This can be partly explained with the fact that Ancient Greek is an inflected language, where meaning is added by means of changing the ending of words rather than adding more words, while English, being only marginally inflected, tends to use more words to convey the same ideas. Moreover, English and Ancient Greek make a very different use of determiners, and English tends to use them much more often, effectively duplicating the number of words used in a given context.

Part of these trends, however, can also be explained as the result of conscious translation choices. The rates of 1-N and N-N links are particularly high for Theodoridis and Johnston: In the former, this substantiates the idea that the translation tends towards expansion, as observed above. Johnston, however, is a close second: Despite the fact that his translation is often semantically similar to the rest of the group, and to Kovacs in particular, the higher 1-N ratio shows a higher tendency towards expansion and paraphrasing. For example, the simple dative of disadvantage “ἀνθρώποις” (lit. “against men”) is translated emphatically as “to lead men astray”; the dative “κακίστῳ”, superlative of the above-mentioned κακόν, with “for a brainless figurine”; the expression “ἐκτίνομεν”, literally “we pay out” (Kovacs) and “we bring to the ground” (Coleridge, more derogatory) is translated as “we must produce a bride price” by Johnston, adding a lot more context than the simple meaning of the word.

To summarize our considerations so far, the combination of various quantitative criteria can be used to reveal certain features of English translations of Ancient Greek. These criteria include: measure of non-aligned words in source and translation; link type ratios; intersection data; part-of-speech intersection. While one single measure can be misleading, the combination of these criteria, coupled with a close examination of the context, can be used to distinguish tendencies to accuracy or, on the contrary, expansion and addition, and to isolate peculiarities in a given text. In the case of Johnston, for example, the intersection data reveals consistent semantic overlap with the other modern and scholarly translation by Kovacs. However, the combined criteria of TP types and non-aligned words suggest that Johnston expands and explains more broadly and freely, and omits more as well. This may be explained as a feature of his translation, which is not meant to be read alongside the original (as opposed to Kovacs), but is conceived for a broad audience of teachers, students, and general readers interested in the ancient world but not necessarily familiar with the language.

Kovacs was expected to be the most consistent translator, according to every measure, and he largely was. However, Coleridge's work shows very similar overall scores, if not even superior (e.g. higher percentage of 1-1 links and lower rate of non-alignment). While the intersection data suggest the peculiarity of the language and distinguishing vocabulary choices of this translator, the other indicators tell a different story, showing how he is still very adherent to the original text, with a very high degree of individual word correspondence, and very little tendency towards significant expansion or omission.

4. *Iliad* 1-67: Comparing Ancient Greek and Persian Translations

The second subset includes alignments of *Iliad* 1.1-67 with three Persian translations:

- Saeed Nafisi, 1958 (Nafisi 1958), derived from the French translation;⁵
- Mir Jalaleddin Kazzazi, 1998. (Kazzazi 1998), derived from the French translation;⁶
- Farnoosh Shamsian, 2020, translated directly from Greek.⁷

All three translations were aligned by the same annotator following the same guidelines. Two out of the three translations are indirect, using French translations as mediating texts. Unlike Greek to English, direct translations from Greek to Persian are rare; however, indirect translation is a common practice and most major texts even have multiple indirect translations, usually from English, French or German. Although indirect translation might be less efficient for translation alignment, it is still the main medium for the transfer of Greek texts to Persian and consequently, has a significant impact on the reception of Greek culture among Persian speakers.

There are similar trends between the two indirect translations in comparison with the direct one. Both indirect translations have a lower number of 1-1 pairs and higher

5. See alignment: <https://www.ugarit.ialigner.com/text.php?id=28503>.

6. See alignment: <https://www.ugarit.ialigner.com/text.php?id=28502>.

7. Available on <https://github.com/farnoosh-shamsian/Iliad>; see alignment: <https://www.ugarit.ialigner.com/text.php?id=28504>.

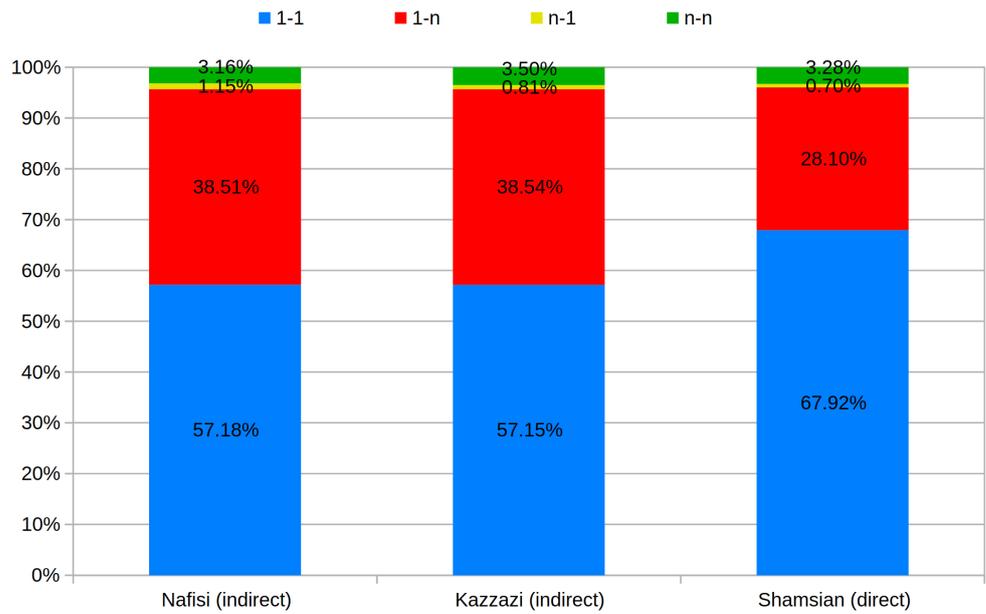


Figure 7: Translation pair ratios across all translations of the *Iliad*.

number of 1-N pairs in comparison to the direct translation. This is mainly caused by phrasal translation of certain Greek words, particularly of epithets. For instance, translation of the word “ἔυκνήμιδες” has 4 tokens both in Kazzazi and in Nafisi, but only 1 token in the direct translation. The ratio of N-N or N-1 pairs doesn’t show considerable differences between the three translations.

However, the most substantial difference is not in the ratio of the pairs, but in the number of non-aligned tokens. The indirect translations are generally longer, Kazzazi’s translation with 901 and Nafisi’s with 742 token in comparison to 542 in the direct translation, and have a much higher number of non-aligned tokens, Nafisi with 233 and Kazzazi with 337, while the non-aligned tokens in the direct translation are minimal.

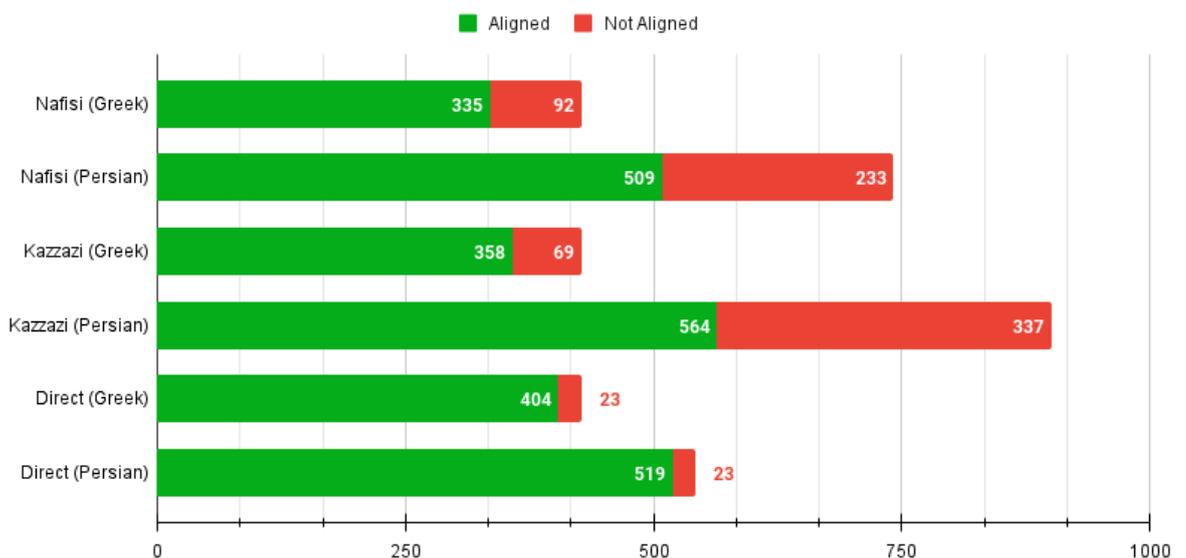


Figure 8: Ratio of aligned and non-aligned tokens across all translations of the *Iliad*.

One reason for the higher number of non-aligned tokens in the indirect translations is that they tend to be more descriptive and use multiple synonyms which have no equivalent in the Greek text but in some cases correspond with the mediating text. Both indirect translations are derived from the French edition by Eugène Lasserre (Homer 1965) while consulting other translations such as Mazon (Homer 1962) and Leconte de Lisle (Homer 1867). The differences might be better demonstrated on a sentence level. For instance, both indirect translations of the *Iliad* 1.25 have 16 tokens and the direct translation has 12 tokens (Except in graphs, all Persian texts have been transcribed for easier formatting).

ἀλλὰ κακῶς ἀφίει , κρατερὸν δ' ἐπὶ μῦθον ἔτελλε (Hom. Il. 1.25)

English translation:

Murray: He sent him away harshly, and laid upon him a stern command

French translations:

Lasserre: Méchamment il renvoya Chrysès, sur cet ordre rude

Mazon: Brutalement il congédie Chrysès, avec rudesse il ordonne

Leconte de Lisle: Il le chassa outrageusement, et il lui dit cette parole violente

Nafisi's translation:

از سر خشم و کبر کریزس را روانه کرد و به سختی و خشونت فرمان داد
 az sar-e khashm va kebr kerīzes rā ravāne kard va be sakhti va khoshūnat farmān dād
 [from] [reason] [anger] [and] [pride] [kerīzes] [rā*] [going] [made] [and] [with] [harshness] [and] [violence] [order] [give]

Kazzazi's translation:

سخت و درشت، شریسه را بازگردانید و خروشان و آتشین خوی، در ستیزه با او، فرمود
 sakht va dorosht, sherīse rā bāzgardānīd va khorūshān o ātashīn-khūy, dar setīze bā ū farmūd
 [harsh][and][coarse][sherīse rā*][sent back][and][roaring][and][with fiery disposition][in][dispute][with][him/her][ordered]

Direct translation:

دگر ظالمانه [خروسس را] معزول می‌داشت و با کلام قاطع فرمان می‌داد
 degar zālemāne (khorūses rā) ma'zūl mīdāsht va bā kalām-e qāṭe' farmān mīdād
 [rather] [ruthlessly] [khoruses] [rā*] [discharged] [was making] [and] [with] [speech] [forceful] [order] [was giving]

*Postposition, commonly used as a marker of the object

Figure 9: Translations of Hom. Il. 1.25 with transcription and glosses.

Since the guideline prioritizes 1-1 alignment, only one of the multiple synonymous equivalents was aligned with the Greek, leaving other synonyms unaligned. In the example of *Iliad* 1.25, the Greek word κακῶς is translated to [zālemāne] in the direct translation, and to [az sar-e khashm va kebr] and [sakht va dorosht] in the indirect translations. The word κρατερὸν in the same line is translated to [qāṭe'] in the direct translation and to [be sakhti va khoshūnat] and [xorūshān va ātashīn-xūy, dar setīze ba 'ū] in the indirect translations.

It should be considered that a change of approach in the guideline could significantly affect the ratio of translation pair. For instance, according to our guidelines, when a word in the Ancient Greek is translated to two or more synonymous word, only one of the equivalents should be aligned. A different approach in the guidelines could have resulted in multiple 1-N pairs by including the synonymous equivalents instead of leaving them unaligned.

Non-aligned tokens of the Greek text Not surprisingly, the number of non-aligned tokens in the Greek text is higher in the indirect translations, 92 in Nafisi, 69 in Kazzazi compared to 23 in the direct translation. Most Greek words without equivalents in the direct translation are particles, often δε and τε, with 8 and 6 incidences respectively out of the 23. On the other hand, the non-aligned tokens in the indirect translations

ἴστων ἐποικομένην (*Hom. Il. 1.31*)

English translation:

Murray: as she walks to and fro before the loom

French translations:

Lasserre: tissant la toile

Mazon: allant et venant devant le métier

Leconte de Lisle: tissant la toile

Nafisi's translation:

آنجا در برابر کارگاه من در رفت و آمد خواهد بود
 ānjā dar barābar-e kārgāh-e man dar raft o āmad khāhad būd
 [there] [in] [front] [workshop] [my] [in] [going] [and] [coming] [will] [be]

Kazzazi's translation:

آنگاه که پیشه‌ای را می‌ورزد
 āngāh ke pīshe'ī rā mīvarzad
 [then] [that] [a profession] [rā*] [practice]

Direct translation:

رجزنان بر دار [بافندگی]
 Rajzanān bar dār-e [bāfandegī]
 [weaving] [on] [loom of] [(weaving)]

*Postposition, commonly used as a marker of the object

Figure 10: Translations of Hom. Il. 1.31 with transcription and glosses.

Translations	Intersection data Iliad 1-67
Nafisi-Kazzazi	70
Nafisi-Shamsian	71
Kazzazi-Shamsian	75
All	70

Table 4: Word intersections across translations of *Iliad*.

also include nouns, verbs and even phrases, caused by semantic variation through the mediating texts.

While the indirect translations do not correspond with the Greek text, they can be aligned with the French translation, particularly with Mazon. For instance, the alignment of Nafisi's translation of Hom.II.1.31 with Mazon would produce the following pairs, leaving only two tokens unaligned in Persian, [ānjā] and [man]:

'allant et venant' - [dar raft-o-āmad khāhad bud], 'devant' - [dar barābar-e], 'métier' - [kārgāh]

Intersections The intersection data extracted from all three translations indicates a high degree of variance and there seems to be no significant difference between direct and indirect translation (see Table 4).

Most of the intersection consists of pronouns and certain particles. Some examples are 'εἴ' - [agar] meaning 'if', 'ἀλλ' - [amā] meaning 'but', 'ἡμῖν' - [mā] meaning 'we', or 'ἐνὶ' - [dar] meaning 'in'. There are also instances of some common words that have a standardized translation, such as 'νυκτὶ' - [shab] meaning night, 'θαλάσσης' - [daryā] meaning sea, or 'πόλεμός' - [jang] meaning war.

Ancient Greek	English	Nafisi	Kazzazi	Direct translation
Ἀχιλλεύς	Achilles	آخيلوس [ākhīlūs]	آشیل [āshīl]	آخیلئوس [ākhīle'ūs]
Ἀτρείδης	Son of Atreus	فرزند آتره [farzand-e ātre]	پور آتره [pūr-e ātre]	پور آترئوس [pūr-e ātre'ūs]
Χρῦσης	Chryses	کریزس [kerīzes]	شریسه [sherīse]	خروسس [khorūses]
Χρῦση	Chryse	کریزه [kerīze]	شریسا [sherīsa]	خروسه [khorūse]

Figure 11: Variations of proper names in translations of the *Iliad*.

Some of these intersections are proper nouns; however, contrary to the high overlap of Greek proper nouns that we see in the English translations, most proper nouns do not match in the Persian translations. Part of these differences is caused by the influence of the mediating language on pronunciation and others by the limits and characteristics of the Persian writing system. Still, few names have only one writing, such as Zeus or Apollo (in Persian, [*āpolon*]). Examples of proper names with multiple spellings are shown in Figure 11.

5. Conclusions and Future Work

In this paper, we presented a preliminary insight into the usage of translation alignment to detect trends in translations of Ancient Greek texts. We presented evidence according to the following criteria: non-aligned words in both languages; intersection, implemented with lemmatization and normalization where possible; frequency of TP types; and part-of-speech data for non-aligned and aligned words in Ancient Greek and English.

We used a combination of these indicators to assess their efficacy in investigating our corpus of translations. For the *Hippolytus*, the analysis led to the somewhat surprising conclusion that a 1910 translation, despite a completely different set of stylistic choices, was in fact closer to the original by most quantitative measures, compared to modern and scholarly ones. For the *Iliad*, the application of the same criteria supported the isolation of phenomena specific to indirect traditions, such as the peculiar rendering of proper names and the use of synonymic expressions as a result of the influx of various French editions. Our study reflects how translators' decisions create very different texts and impressions of the original: While lack of vocabulary overlap is to be expected, it is very surprising to see how little consistency there is in addressing even the least ambiguous words. Overall, patterns of disagreement can be detected across the board, and make certain translations emerge for their peculiarities.

The work here presented is part of a larger effort in upscaling the functionalities of Ugarit: It is conducted in parallel with the development of alignment guidelines for various types of research, and with the implementation of an automatic alignment model for Ancient Greek (Yousef et al. 2022c, Yousef et al. 2022a). The implementation

of an automatic model should considerably alleviate the burden of the manual work required for a study like this, and allow for the expansion of the corpus. In turn, a larger corpus should increase the threshold of tolerance to errors, which will be crucial with larger corpora, where tasks of massive normalization and POS tagging will be necessary.

However, in the course of this study we have also observed that the manual intervention of a scholar in the establishment of certain kinds of links is essential to conduct an analytical study. Automatic models tend to privilege 1-1 TPs, but a researcher may be more interested in investigating phrasal translations, or collecting instances of peculiar expansion, and so on. This implies that there always needs to be a certain level of close reading, regardless of corpus size, and that alignment guidelines need to be designed with certain research questions in mind. Finally, it needs to be emphasized that some languages, such as Persian, do not have sufficient support in NLP architecture to allow for extensive automatic analysis. Nevertheless, it is extremely important that this type of research is pursued, and this further shows the demand of support for low-resourced languages in the digital space.

Traduttore traditore.⁸ Translations may appear equivalent on the surface, but they are really different in the way they render the complexities of an ancient text. As their semantic overlap is minimal, they reflect the individuality of the translators and their circumstances, rather than just the character of the author. However, translations are necessary, and can even be works of art in their own respect. As computational methods become more accessible and advanced, tools like Ugarit can empower an in-depth approach to the study of texts, by facilitating a user-centered and controlled approach to word alignment. The application of translation alignment as a pedagogical and scholarly method can empower a cross-linguistic attitude to the text, where the focus is not only on the translation or on the original, but on the meaningful exchange occurring between them, at the semantic, grammatical, and conceptual level.

6. Data Availability

- Repository: <https://github.com/UgaritAlignment/JCLS22-Paper>;
- *Hippolytus* pt.1: <https://ugarit.ialigner.com/userProfile.php?userid=3&tid=9964>;
- *Hippolytus* pt.2: <https://ugarit.ialigner.com/userProfile.php?userid=3&tid=9965>;
- *Iliad*, Nafisi: <http://www.ugarit.ialigner.com/text.php?id=28503>;
- *Iliad*, Kazzazzi: <http://www.ugarit.ialigner.com/text.php?id=28502>;
- *Iliad*, Shamsian: <http://www.ugarit.ialigner.com/text.php?id=28504>.

8. Italian for “translator traitor”.

7. Software Availability

The alignments have been created using UGARIT translation alignment editor
<http://ugarit.ialigner.com>

8. Acknowledgements

We would like to thank Bethany Morgan for providing the first version of the alignments of *Hippolytus*, for selecting the translations, and for her valuable feedback on this article; Gregory Crane and Monica Berti for their guidance, as always. Additional thanks to all the participants of the CCLS2022 Conference in Darmstadt, for their enthusiastic reception and thoughtful questions.

9. Author Contributions

Chiara Palladino: Conceptualization, Formal Analysis, Investigation, Writing-original draft, Validation

Farnoosh Shamsian: Writing-original draft, Formal analysis, Investigation, Resources

Tariq Yousef: Methodology, Software, Data curation, Visualization, Review & Editing

References

- Almas, Bridget and Marie-Claire Beaulieu (2013). “Developing a New Integrated Editing Platform for Source Documents in Classics”. In: *Literary and Linguistic Computing* 4 (28), 493–503. [10.1093/lc/fqt046](https://doi.org/10.1093/lc/fqt046).
- Baker, Mona (1993). “Corpus Linguistics and Translation Studies: Implications and Applications”. In: *Text and Technology: In Honour of John Sinclair*. Ed. by Mona Baker, Gill Francis, and Elena Tognini-Bonelli. John Benjamins, 233–250. [10.1075/z.64.15 bak](https://doi.org/10.1075/z.64.15bak).
- Barreiro, Anabela, Francisco Raposo, and Tiago Luís (2016). “CLUE-Aligner: An Alignment Tool to Annotate Pairs of Paraphrastic and Translation Units”. In: *Proceedings of the LREC 2016 Workshop “Translation Evaluation – From Fragmented Tools and Data Sets to an Integrated Ecosystem”*. Ed. by Georg Rehm, Aljoscha Burchardt, Ondřej Bojar, Christian Dugast, Marcello Federico, Josef van Genabith, Barry Haddow, Jan Hajič, Kim Harris, Philipp Koehn, Matteo Negri, Martin Popel, Lucia Specia, Marco Turchi, and Hans Uszkoreit, 7–13. http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-MT%20Evaluation_Proceedings.pdf (visited on 11/21/2022).
- Bettini, Maurizio (2012). *Vertere: un’antropologia della traduzione nella cultura antica*. Einaudi.
- Bizzoni, Yuri, Marianne Reboul, and Angelo Del Grosso (2017). “Diachronic Trends in Homeric Translations”. In: *Digital Humanities Quarterly* 11 (2). <http://www.digitalhumanities.org/dhq/vol/11/2/000297/000297.html> (visited on 11/21/2022).

- Caseli, Helena de Medeiros, Valéria Delisandra Feltrim, and Maria das Graças Volpe Nunes (2002). *TagAlign: Uma ferramenta de pré-processamento de textos (NILC-TR-02-09)*. Tech. rep. 169. <http://www.nilc.icmc.usp.br/nilc/download/NILC-TR-02-09.zip> (visited on 11/21/2022).
- Crane, Gregory, Neven Jovanovic, Sophia Sklaviadis, Margherita de Luca, Petra Šoštarić, Maryam Foradi, Kate Cottrell, James Tauber, Farnoosh Shamsian, and Chiara Palladino (2019). “Confronting Complexity of Babel in a Global and Digital Age. How Can You Work with a Language that You Do Not Know?” In: *Book of Abstracts of the Digital Humanities Conference 2019*. ADHO. <https://dh-abstracts.library.virginia.edu/works/9793> (visited on 11/21/2022).
- Dagan, Ido, Kenneth Church, and William Gale (1999). “Robust Bilingual Word Alignment for Machine Aided Translation”. In: *Natural Language Processing Using Very Large Corpora*. Ed. by Susan Armstrong, Kenneth Church, Pierre Isabelle, Sandra Manzi, Evelyne Tzoukermann, and David Yarowsky. Text, Speech and Language Technology. Springer, 209–224. [10.1007/978-94-017-2390-9_13](https://doi.org/10.1007/978-94-017-2390-9_13).
- Dou, Zi-Yi and Graham Neubig (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*. Association for Computational Linguistics, 2112–2128. [10.18653/v1/2021.eacl-main.181](https://doi.org/10.18653/v1/2021.eacl-main.181).
- Germann, Ulrich (2008). “Yawat: Yet Another Word Alignment Tool”. In: *Proceedings of the ACL-08: HLT demo session*, 20–23. <https://aclanthology.org/P08-4006/> (visited on 11/21/2022).
- Gibert, John (2022). “Review of: Euripides, Children of Heracles, Hippolytus, Andromache, Hecuba”. In: *Bryn Mawr Classical Review* (Bmcr Id: 1996.12.02). <https://bmcrr.brynmawr.edu/1996/1996.12.02/> (visited on 11/21/2022).
- Gilmanov, Timur, Olga Scriver, and Sandra Kübler (2014). “SWIFT Aligner, A Multifunctional Tool for Parallel Corpora: Visualization, Word Alignment, and (Morpho)-Syntactic Cross-Language Transfer.” In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, 2913–2919. [10.18653/v1/2021.eacl-main.181](https://doi.org/10.18653/v1/2021.eacl-main.181).
- Graça, João, Joana Paulo Pardal, Luísa Coheur, and Diamantino Caseiro (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)*, 986–993. http://www.lrec-conf.org/proceedings/lrec2008/pdf/250_paper.pdf (visited on 11/21/2022).
- Grimes, Stephen, Xuansong Li, Ann Bies, S. Kulick, Mam Xiaoyi, and Stephanie Strassel (2010). “Creating Arabic-English Parallel Word-Aligned Treebank Corpora at LDC”. In: *Proceedings of Language Resources and Evaluation Conference (LREC’10)*. ELRA. https://catalog.ldc.upenn.edu/docs/LDC2019T18/LREC2010_Arabic_parallel_aligned_TB.pdf (visited on 11/21/2022).
- Homer (1867). *Homère: Iliade ; traduction nouvelle par Leconte de Lisle*. Alphonse Lemerre.
- (1962). *Iliade: Traduction de Paul Mazon*. Les Belles Lettres.
- (1965). *L’Iliade: traduction, introduction et notes par Eugène Lasserre*. Garnier-Flammarion.
- Jalili Sabet, Masoud, Philipp Dufter, François Yvon, and Hinrich Schütze (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*. Association for Computational Linguistics, 1627–1643. [10.18653/v1/2020.findings-emnlp.147](https://doi.org/10.18653/v1/2020.findings-emnlp.147).
- Kay, Martin and Martin Röscheisen (1993). “Text-translation Alignment”. In: *Computational Linguistics* 19 (1), 121–142. <http://dl.acm.org/citation.cfm?id=972450.972457> (visited on 11/21/2022).
- Kazzazi, Mir Jalaleddin (1998). *Iliad*. Markaz.
- Laviosa, Sara (2008). “Corpus-based Translation Studies: Where does it come from? Where is it Going?” In: *Language Matters* 35 (1). [10.1080/10228190408566201](https://doi.org/10.1080/10228190408566201).
- Melamed, I. Dan (1998). “Manual Annotation of Translational Equivalence: The Blinker Project”. In: *arXiv preprint*. [10.48550/arXiv.cmp-lg/9805005](https://arxiv.org/abs/10.48550/arXiv.cmp-lg/9805005).
- Munson, Rosaria V. (2001). “Ananke in Herodotus”. In: *The Journal of Hellenic Studies* 121, 30–50. [10.2307/631826](https://doi.org/10.2307/631826).
- Nafisi, Saeed (1958). *Iliad*. Elmi Farhangi.
- Nergaard, Siri (1993). *La teoria della traduzione nella storia*. Bompiani.
- Och, Franz Josef and Hermann Ney (2003). “A Systematic Comparison of Various Statistical Alignment Models”. In: *Computational Linguistics* 29 (1), 19–51. [10.1162/089120103321337421](https://doi.org/10.1162/089120103321337421).
- Palladino, Chiara (2020). “Reading Texts in Digital Environments: Applications of Translation Alignment for Classical Language Learning”. In: *The Journal of Interactive Technology and Pedagogy* 18. <https://jitp.commons.gc.cuny.edu/reading-texts-in-digital-environments-applications-of-translation-alignment-for-classical-language-learning/> (visited on 11/21/2022).
- Palladino, Chiara, Maryam Foradi, and Tariq Yousef (2021). “Translation Alignment for Historical Language Learning: a Case Study”. In: *Digital Humanities Quarterly* 15 (3). <http://digitalhumanities.org/dhq/vol/15/3/000563/000563.html> (visited on 11/21/2022).
- Shukhoskvili, Maia (2017). “Methodology of Translation Alignment of Georgian Text of Plato’s “Theaetetus””. In: *International Journal of Language and Linguistics* 4 (4), 63–69. <https://www.ijllnet.com/journal/index/2393> (visited on 11/21/2022).
- Véronis, Jean, ed. (2000). *Parallel Text Processing: Alignment and Use of Translation Corpora*. Text, Speech and Language Technology. Springer.
- Yousef, Tariq, 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 13th Conference on Language Resources and Evaluation (LREC 2022)*, ELRA, 5894–5905. <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.634.pdf> (visited on 11/21/2022).
- Yousef, Tariq, Chiara Palladino, Farnoosh Shamsian, and Maryam Foradi (2022b). “Translation Alignment with Ugarit”. In: *Information* 2 (13). [10.3390/info13020065](https://doi.org/10.3390/info13020065).
- Yousef, Tariq, Chiara Palladino, David J. Wright, and Monica Berti (2022c). “Automatic Translation Alignment for Ancient Greek and Latin”. In: *OSF Preprints*. [10.31219/osf.io/8epsy](https://doi.org/10.31219/osf.io/8epsy).