

Translation Alignment for Historical Language Learning: a Case Study

Chiara Palladino <chiara_dot_palladino_at_furman_dot_edu>, Classics, Furman University
Maryam Foradi <maryam_dot_foradi_at_uni-leipzig_dot_de>, University of Leipzig
Tariq Yousef <tariq_dot_yousef_at_uni-leipzig_dot_de>, Institute of Computer Science, University of Leipzig

Abstract

This paper proposes text alignment in digital environments as a way to empower language learning. It presents the principles and goals of text alignment in Natural Language Processing, and introduces Ugarit, a web-based translation alignment editor for the collection of aligned language pairs. Then, it reports observations on the application of translation alignment in historical language courses at Tufts and Furman University between 2017 and 2019.

Introduction

The World Wide Web has introduced a massive change in the cognitive processes of reading: skimming has become the most popular way in which readers interact with written texts on digital support, with possible enormous consequences on how human beings process information and articulate complex thought [Wolf 2018]. On the other hand, digital environments have an immense, often underdeveloped, potential to improve interactive engagement and exploration of information, integrating very different media and conveying extremely complex and multifaceted messages. Readers of the future should be educated to critically approach a text through digital technologies, making full use of this potential, and going beyond the reading processes conceived for texts encoded on static supports. In this paper, we propose the application of text alignment as one method to empower the reflective engagement with digital texts, specifically in the context of language learning.

1

Text alignment, and more specifically translation alignment, is a type of annotation,^[1] which consists in analytically matching a word or expression in a text with the corresponding ones in another text, often in a different language, in environments created to collect aligned word pairs. This is an exceptionally complex hermeneutic task, by which readers are induced to problematize the function and implications of morphosyntactic, semantic, and expressive devices of a language, even of their own native tongue.

2

This paper collects a set of preliminary observations on groups of students that have used translation alignment in the framework of language courses at various levels of expertise. We propose text alignment as a way to empower the perception of the complexity of a written source, but also as a method to leverage usual obstacles in the process of reading in a different language by directly engaging with original literary texts.

3

Principles of Translation Alignment

Translation alignment is one of the most popular applications of Natural Language Processing. It is defined as the comparison of two or more texts in different languages, also called parallel texts or parallel corpora [Kay and Röscheisen 1993] [Véronis 2000],^[2] by means of automated or semi-automated methods. The main purpose is to define which parts in a source text correspond to which parts of a second text in a different language. The result often takes the form of a list of pairs of items, which can be words, sentences, or larger text chunks like paragraphs or documents. This task is often performed automatically through various methods, which often require large amounts of

4

training data: one of the most traditional methods for automated translation alignment is statistical machine translation [Brown et al. 1990], which uses statistical methods to extract and identify correspondences of words and phrases in two parallel texts, based on already available aligned pairs.

The alignment of texts in different languages, however, is an exceptionally complex task, because of the several variables involved. It is often difficult to find perfect correspondences across languages that express ideas through different morphosyntactic constructs, with variations in word order, sentence length, and even underlying cultural significance. Machine-actionable systems are often inefficient in providing equivalences for wordplays, metaphors, and other rhetorical devices. The resulting aligned pairs may be one-to-one (one word in the source text corresponds to one word in the translation), but often align as one-to-many, many-to-many, or many-to-one. Each word correspondence may be complete or perfect (with complete overlap between two words), but also possible or incomplete (partial overlap, or both words being a translation of each other only in certain contexts [Graça et al. 2009]).

For these reasons, manually aligned word pairs are one of the main sources of training data, and are not only employed for the implementation of machine translation systems, but are increasingly useful for many other purposes, including information extraction and automated creation of bilingual lexica [Dagan et al. 1999] [Melamed 1998]. Therefore, large corpora of manually aligned word pairs are a much desired resource, which remains, however, extremely rare.

The operation of collecting training data is often configured as a Citizen Science effort: for example, Glosbe (<https://glosbe.com/>) provides thousands of dictionaries created from the manual alignment of word pairs performed by a community of users.^[3] However, because of the hermeneutic complexity implied in the alignment across different languages, translation alignment has also a powerful pedagogical potential. We have implemented a specific web-based environment, called Ugarit, to provide a framework for language learning applications through translation alignment.

The Ugarit Translation Alignment Editor

The first public web-based environment for manual text alignment of parallel corpora was conceived within the Alpheios Project (<http://alpheios.net/>), and is currently hosted by the Perseids Project (<https://www.perseids.org/>). The editor offers an easy environment to manually match corresponding words in two parallel texts.

The Ugarit Translation Alignment Editor (<http://ugarit.ialigner.com/>), developed in 2016 by Tariq Yousef at the University of Leipzig, was conceived as the first version of an improved interface that could perform more extensive alignment tasks than the Alpheios editor: it allows alignment in up to three texts, incorporates transliteration of non-Latin alphabets, and collects manually aligned pairs as training datasets [Yousef 2019] [Yousef and Jänicke 2020]. Moreover, it allows almost immediate visible online publication of the material.

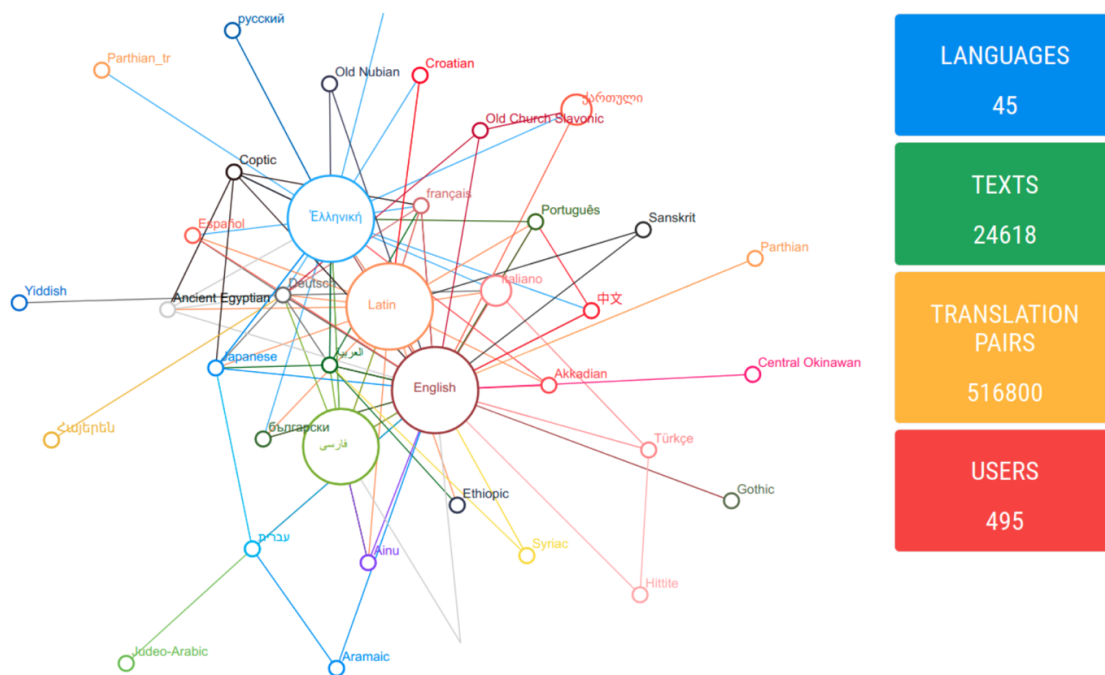
The first version of the tool was made public in March 2017. Since then, Ugarit has registered an ever-increasing number of visits and new users.

Originally, Ugarit was developed to collect training data for the implementation of statistical machine translation of historical languages, mainly Ancient Greek, Latin, and Persian, for which few to none aligned datasets exist. Ideally, historical languages are closed systems with a finite number of words and very limited change in the foreseeable future. Therefore, it should be possible to create adequately efficient automated methods of statistical machine translation based on a relatively small training dataset.

However, after the tool was made public, the number and variety of languages included by the users has steadily increased and has gone far beyond the original intent: at the moment this article is being written, 45 languages are included in Ugarit, including Ancient Greek, Latin, Arabic, Persian, English, French, German, Italian, Portuguese, Croatian, Armenian, Japanese, Chinese, Syriac, Sanskrit, Coptic, Egyptian, Akkadian, and Ethiopic; 465 unique users, and about 24,500 parallel texts, are currently hosted.

The home page of Ugarit offers a quick overview of the languages currently available (the graph showing the

proportions of the respective corpora), the number of users and texts, the available word pairs, and the newest alignments.



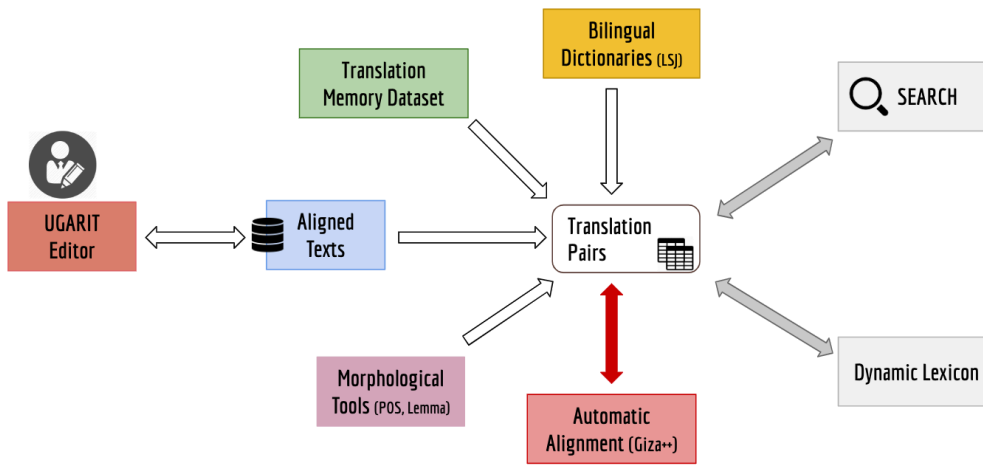
The workflow is simple: users can upload their corpus in text format, or import it by calling its CTS URN from the Perseus Digital Library (<http://www.perseus.tufts.edu/hopper/>) [Babeu 2019]. Then, the parallel texts can be manually aligned by matching words or groups of words. Resulting pairs are automatically stored in the database, and can then be exported in XML or tabular format. 14

Currently, Ugarit works with up to three texts in different languages, and there is virtually no limit to the amount and length of texts that can be uploaded. A clickable option enables the user to decide whether the alignment can be publicly visible on the website or not. When the default value is kept, the text appears in the homepage of the website, and other users can inspect the alignment by hovering with the mouse on individually paired tokens, which are automatically highlighted. The proportion of aligned tokens is indicated in the colored bar below the text: the green indicates the rate of matching words, the red the rate of non-matching words. 15

For languages with non-Latin alphabets, Ugarit offers automatic transliteration, which is visible when the pointer hovers on the desired word. This feature is currently available for Greek, Arabic, Persian,^[4] Armenian, and Georgian. 16

<p>Հայերեն</p> <p>Եւ կայսր Ժամանակս ընդ այնտիկ խնդրեաց իւր կին կայսրն Դիոկղետիանոս .</p> <p>(12) 100% HYE - GRC</p>	<p>Ἑλληνική</p> <p>Καὶ ἐγένετο ἐν ἐκείνοις τοῖς καιροῖς , ἐζήτησεν ἑαυτῷ γυναῖκα Διοκλητιανός ὁ Καίσαρ πρὸς γάμου κοινωνίαν .</p> <p>(16) 94% HYE - GRC (1)</p>	<p>Latin</p> <p>Illis temporibus Diocletianus imperator quæsitit mulierem nuptiis sibi sociandam :</p> <p>(16) 94% HYE - GRC (1)</p>
---	---	--

Translation pairs are stored in a local database that gathers data from the manually aligned texts, but also from bilingual dictionaries and automatically aligned texts with GIZA++.^[5] This database would form the bulk of training data to improve automatic translation alignment methods focusing on historical languages. 17



As such, the database of translation pairs serves two main functions: first, it serves search queries to provide information on all the available translations of a word. Therefore, if the user looks for a word (for instance, “life”) with the search function available on the home page, the output displays all the available aligned pairs currently hosted in Ugarit. In this way, the user can access an extended number of translations of a word in many different languages.

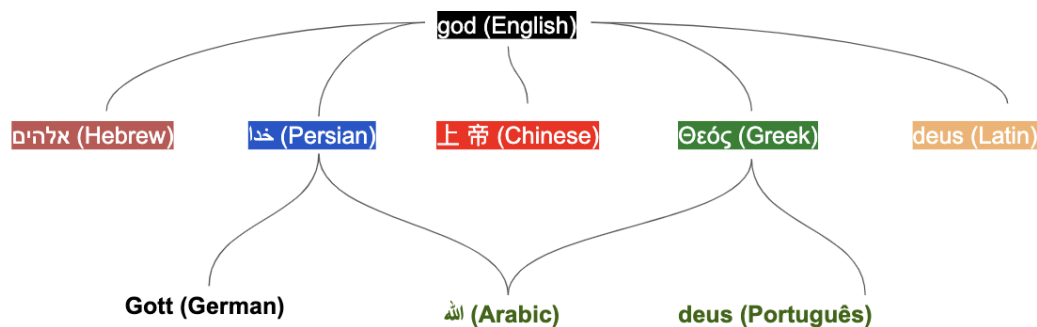
English

616 results are found for *life*

Frequency: 859

Ελληνική	βίον 41 βίου 28 ψυχὴν 16 ζῆν 16 βίος 10 ζωὴ 10 ψυχῆς 7 σώματος 6 βίῳ 3 ζωῆς 2 αἰῶνος 2 ἀποκτεῖναι 2 βίωτος 3 βίωτον 3 ζῆ 2 βεβιωκότας 2 σώματι 2 σωθῆναι 2 ζῶν 2 ὦν 2 αἰῶνα 2 ἀπολαυστικός 1 τρόποι 1 ἀφ᾽εἰ 1 τὸν βίον 1 ἀποκτείνῃ 1 ἀποθανόντι 1 σωτηρία 1 ζωῆ 1 οὐδ᾽ 1 ἀπέθανον 1 ψυχῇ 1 διαίταν 1 κεφαλῇ 1 βίους 1 ἀπολοιμν 1 ζυμῶσιον 1 διαίτης 1 ἀποθανεῖν 1 διάγουσι 1 ζῶντι 1 δυνατοί 1 εἶσοε 1 δεφεστῆρι 1 ἐνδειαν 1 βίων 1 ζωῆν 1 ἀπολέσει 1 αἰῶν 1 ἀποκτείναντας 1 θάλλει 1 τὸν πότμον 1 φιλόμοχοι 1 φιλόμοστον 1 θυμὸν 1 ἐξηρασμένα 1 δέμας 1 βιώτους 1 ψυχῇ 1 ἐζης 1 αἰῶνι 1 ψυχῆν 1 βίωτος 1 ζῶα 1 βεβιωκότι 1 λειπόμενον 1 διατριβῶν 1 ἐπιτήδευμα 1 βεβιωκέναι 1 ζῶντες 1 φωνῆος 1 θανάτω 1 σώματα 1 σφρασε 1 ῥησυχίαν 1 ἐπιτήδευσμάτων 1 θάνατον 1 ζῶντος 1 πολιτείας 1 τοῦ βίου 1
English	life 2 his life 1
中文	人生 2 寿 1
فارسی	جان 31 عمر 31 حیات 7 حیوان 7 زندگی 3 زندگی 7 سن 1 سر 1 روان 1 حیوانات 1 خون 1 محیا 1 عمرت 1 خضر 1
Latin	vita 3 vitae 2 vitam 1 vivo 1
Português	vida 4 a vida 2
Ancient Egyptian	'nh 2

Second, Ugarit provides training data to the Dynamic Lexicon (<http://dynamiclexicon.com/>), which applies the principle of triangulation to extract bilingual pairs across Ancient Greek, Persian, and Latin, using bridge pairs in Greek/English, Persian/English, and Latin/English: the Dynamic Lexicon fetches alignment information from all available aligned corpora, and it uses aligned languages as pivot to fetch translated words in other languages that may bear the same meaning [Berti and Yousef 2015]. So, each aligned pair functions as a translation vector to access more matches in other languages that are not directly aligned. This method is based on the assumption that two words in two given languages are more likely to match if they are translations of the same word in a third language [Kumar et al. 2007]. The accuracy of the triangulation can be improved by expanding the available corpus of translation pairs, considering available part-of-speech classification, or adding more pivot languages [Yousef 2015].



In other words, Ugarit was designed as a Citizen Science resource to collect training datasets on historical languages from a variety of sources and projects. One of the most important initiatives in this regard was the Hafez Project (<http://dynamiclexicon.com/hafez/>; <http://www.divan-hafez.com>) led by Maryam Foradi at the University of Leipzig [Foradi et al. 2017], which tested how German and Persian speakers aligned textual portions of Hafez using English as a bridge language. The results indicated that, provided with the appropriate scaffolding by using English as a bridge language, users with no knowledge of the source language could generate word alignments with the same output accuracy generated by expert users in both languages of source and target texts. The study showed that, while the long-term pedagogical effectiveness of this method is comparable with learning vocabulary with flashcards, the vocabularies learned both intentionally with flashcards and incidentally through text alignment had a significantly better retention [Foradi 2019].

20

These results suggested that alignment could serve as a pedagogical tool with a certain effect of long-term retention of vocabulary. Between 2017 and 2019, we conducted a pilot study investigating how students of historical languages could use translation alignment to improve their learning experience through direct engagement with original texts. The lack of scholarship on the pedagogical application of alignment as a tool for language learning, and the complexity involved in the reproducibility of the experience, compel us to start with pre-experimental observations from specific cases, where translation alignment has been applied as a standard method during language courses. Our initial observations provide the justification for a more experimental design.

21

Our Pilot Study

Ugarit was used in graduate and undergraduate semester-long Classics courses at the Universities of Tufts and Furman, with different focuses: language (mainly Ancient Greek and Latin) at elementary, intermediate, and upper level, literature in translation, literature surveys, and individual research projects.^[6] The use of translation alignment was incorporated in the teaching practice as a structural element:

22

- The students made continuous use of Ugarit in class, using and creating aligned translations as a reading support.
- The task of aligning more lengthy chosen texts against one or two translations constituted the main type of assignment, in the form of individual research projects at various degrees of complexity: depending on the course and level, the students produced bilingual and trilingual alignments, whose accuracy was graded accordingly.
- The results were discussed in the form of research papers and presentations, in class and also in public venues.^[7]

The students reported their observations on the process, evaluated how it affected their understanding of the source, and performed a qualitative analysis on the languages in question, examining specific morphosyntactic, semantic, and expressive phenomena.

We observed the application of this workflow on students with very diverse language skills, including those who mastered other modern languages, and those whose native language was not English. We propose three cases of study:

23

1. Students who were proficient in the source language and had complete mastery of one other language (including the native tongue).
2. Students who were proficient in two different languages, and had basic knowledge of a third one.
3. Students who were proficient in one or more languages, but had no prior knowledge of the source language.

In the following paragraphs, we report the considerations formulated by the students themselves, and our observations on the outcomes of the process.

Case 1: Students of the source language

Students of Case 1 were enrolled in intermediate and upper level courses in historical languages, mainly Ancient Greek or Latin. They performed two tasks:

24

1. Pairwise alignment between the source text and one chosen translation in a modern language.
2. Alignment of the source text against two different translations.

The purpose of the first exercise was to set the groundwork for a structural analysis of the most important morphosyntactic and semantic devices of the source language, comparing how those features were rendered in a different language (often English), of which the students had complete mastery. In the second exercise, they built upon this experience to draw an evaluation of two competing translations, focusing on the different ways in which complex linguistic phenomena can be rendered in another language.

Example 1: Plato, *Symposium* 210-212. Ancient Greek-English.^[8]

Ελληνικά	Transliterate	English
<p>ὀρθῶς ἂν μέχρι ἐνταῦθα πρὸς τὰ ἐρωτικά παιδαγωγῆθῃ, θεώμενος ἐφεξῆς τε καὶ ὀρθῶς τὰ καλά, πρὸς τέλος ἦδη ἰῶν τῶν ἐρωτικῶν ἐξαίφνης κατόψεται τι θαυμαστόν τῆν φύσιν καλόν, τοῦτο ἐκείνο, ὃ Σώκρατες, οὐ δὴ ἔνεκεν καὶ οἱ ἔμπροσθεν πάντες πόνοι ἦσαν, πρῶτον μὲν αἰεὶ ὄν καὶ οὔτε γινόμενον οὔτε ἀπολλύμενον, οὔτε αὐξανόμενον οὔτε φθίνον, ἔπειτα οὐ τῆ μὲν καλόν, τῆ δ' αἰσχρόν, οὐδὲ τὸ μὲν, τοτὲ δὲ οὐ, οὐδὲ πρὸς μὲν τὸ καλόν, πρὸς δὲ τὸ αἰσχρόν, οὐδ' ἔνθα μὲν καλόν, ἔνθα δὲ αἰσχρόν, ὡς τίσι μὲν ὄν καλόν, τίσι δὲ αἰσχρόν: οὐδ' αὖ φαντασθήσεται αὐτῷ τὸ καλὸν οἷον πρόσωπόν τι οὐδὲ χεῖρας οὐδὲ ἄλλο οὐδὲν ἂν σῶμα μετέχει, οὐδὲ τις λόγος οὐδὲ τις ἐπιστήμη, οὐδὲ που ὄν ἐν τέρψῃ τι, οἷον ἐν ζῶψῃ ἢ ἐν γῆ ἢ ἐν οὐρανῷ ἢ ἐν τῷ ἄλλῳ, ἀλλ' αὐτὸ καθ' αὐτὸ μὲθ' αὐτοῦ μονοειδὲς αἰεὶ ὄν, τὰ δὲ ἄλλα πάντα καλά ἐκείνου μετέχοντα τρόπον τινὰ τοιοῦτον, οἷον γινόμενον τε τῶν ἄλλων καὶ ἀπολλυμένων μὴδὲν ἐκείνου μήτε τι πλέον μήτε ἔλαττον γίνεσθαι μὴδὲ πάσχειν μὴδέν. ὅταν δὴ τις ἀπὸ τῶνδε διὰ τὸ ὀρθῶς παιδραστεῖν ἐπιανῶν ἐκείνου τὸ καλὸν ἀρχηται καθορᾶν, σχεδὸν ἂν τι ἅπτοιο τοῦ τέλους. τοῦτο γὰρ δὴ ἐστὶ τὸ ὀρθῶς ἐπὶ τὰ ἐρωτικά ἵναί ἢ ὑπ' ἄλλου ἀγεσθαι, ἀρχόμενον ἀπὸ τῶνδε τῶν καλῶν ἐκείνου ἔνεκα τοῦ καλοῦ αἰεὶ ἐπιανέιναι, ὡσπερ ἐπιαναστασίς χρώμενον, ἀπὸ ἐνός ἐπὶ δύο καὶ ἀπὸ δύοιν ἐπὶ πάντα τὰ καλά σῶματα, καὶ ἀπὸ τῶν καλῶν σωματίων ἐπὶ τὰ καλά ἐπιτηδεύματα, καὶ ἀπὸ τῶν ἐπιτηδεύματων ἐπὶ τὰ καλά μαθήματα, καὶ ἀπὸ τῶν μαθημάτων ἐπ' ἐκείνο τὸ μάθημα τελευτήσῃ, ὃ ἐστὶν οὐκ ἄλλου ἢ αὐτοῦ ἐκείνου τοῦ καλοῦ μάθημα, καὶ γνῶ αὐτὸ τελευτῶν ὃ ἐστὶ καλόν. ἐνταῦθα τοῦ βίου, ὃ φιλεῖ Σώκρατες, ἔφη ἡ Μαντινικὴ ζῆνη, εἰπερ που ἄλλοθι, βιωτῶν ἀνθρώπων, θεωμένῳ αὐτὸ τὸ καλόν. ὃ ἐάν ποτε ἴδῃς, οὐ κατὰ χρυσίον τε καὶ ἐσθῆτα καὶ τοὺς καλοὺς παιδάς τε καὶ νεανίσκους δόξει σοὶ εἶναι, οὐς νῦν ὄρων ἐκπέπληξαι καὶ ἔταμος εἶ καὶ οὐ καὶ ἄλλοι πολλοί, ὀρώντες τὰ παιδικὰ καὶ συνόντες αἰεὶ αὐτοῖς, εἰ πως οἷόν τ' ἦν, μήτ' ἐσθίειν μήτε πίνειν, ἀλλὰ θεασθαι μόνον καὶ συνεῖναι. τί δῆτα, ἔφη, οἰόμεθα, εἰ τῷ γένειτο αὐτὸ τὸ καλόν ἰδεῖν εὐκρινές, καθαρὸν, ἄμεικτον, ἀλλὰ μὴ ἀνάμλεον σαρκῶν τε ἀνθρωπίνων καὶ χρωμάτων καὶ ἄλλης πολλῆς φουαρίας θνητῆς, ἀλλ' αὐτὸ τὸ θεῖον καλὸν δύναιτο μονοειδὲς κατιδεῖν; ἄρ' οἶε, ἔφη, φαῦλον βίον γίνεσθαι ἐκεῖσε βλέποντος ἀνθρώπου καὶ ἐκείνο ᾧ δεῖ θεωμένου καὶ συνόντος αὐτῷ; ἢ οὐκ ἐνθυμῆ, ἔφη,</p>		<p>"You see, the man who has been thus far guided in matters of Love, who has beheld beautiful things in the right order and correctly, is coming now to the goal of Loving: all of a sudden he will catch sight of something wonderfully beautiful in its nature; that, Socrates, is the reason for all his earlier labors: First, it always is and neither comes to be nor passes away, neither waxes nor wanes. Second, it is not beautiful this way and ugly that way, nor beautiful at one time and ugly at another, nor beautiful in relation to one thing and ugly in relation to another; nor is it beautiful here but ugly there, as it would be if it were beautiful for some people and ugly for others. Nor will the beautiful appear to him in the guise of a face or hands or anything else that belongs to the body. It will not appear to him as a speech or a kind of knowledge does. It is not anywhere in another thing, as in an animal, or in earth, or in heaven, or in anything else, but itself by itself with itself, it is always one in form; and all the other beautiful things share in that, in such a way that when those others come to be or pass away, this does not become the least bit smaller or greater nor suffer any change. So when someone rises by these stages, through loving boys correctly, and begins to see this beauty, he has almost grasped his goal. This is what it is to go aright, or be led by another, into the art of Love: one goes always upwards for the sake of this Beauty, starting out from beautiful things and using them like rising stairs: from one body to two and from two to all beautiful bodies, then from beautiful bodies to beautiful customs, and from customs to learning beautiful things, and from these lessons he arrives in the end at this lesson, which is learning of this very Beauty, so that in the end he comes to know just what it is to be beautiful. And there in life, Socrates, my friend," said the woman from Mantinea, "there if anywhere should a person live his life, beholding that Beauty. If you once see that, it won't occur to you to measure beauty by gold or clothing or beautiful boys and youths — who, if you see them now, strike you out of your senses, and make you, you and many others, eager to be with the boys you love and look at them forever, if there were any way to do that, forgetting food and drink, everything but looking at them and being with them. But how would it be, in our view," she said, "if someone got to see the Beautiful itself, absolute, pure, unmixed, not polluted by human flesh or colors or any other great nonsense of</p>

Graduate, native English speaker, upper level Ancient Greek. The student created bilingual alignments of Plato's *Symposium*, comparing it against the translation by Nehamas and Woodruff (1989), in the context of a research project^[9] that integrated different levels of linguistic digital annotation, with the goal of enabling readers with no Greek knowledge to explore the original through the dynamic reading of the aligned parallel texts. The student created a consistent alignment strategy that only matched words corresponding in meaning and grammatical function (with the partial exception of verbs for which a perfect match would not be possible, such as subjunctives or participles with predicative use).

25

Example 2: Apuleius, *Metamorphoses* 8.19-22. Latin-English.^[10]

<p>Latin</p> <p>Interea quidam senex de summo colle prospectat , quem circum capellae pascentes opilionem esse profecto clamabant . Eum rogavit unus e nostris habere tene veni lactem vel adhuc liquidum vel in caseum recentem inchoatum . At ille diu capite quassanti 'Vos autem ' inquit 'De cibo vel poculo) vel omnino ulla refectione nunc cogitatis ? An nulli scitis quo loco comederitis ? ' Et cum dicto conductis ovis conversus longe recessit . Quae vox eius et fuga pastoribus nostris non mediocrem pavorem incussit : ac dum perterriti de loci qualitate sciscitari gestiunt nec est qui doceat , senex alius , magnus ille quidem , gravatus annis , totus in baculum pronus et lassum trahens vestigium , ubertim lacrimans per viam proximat , visisque nobis cum fletu maximo singulorum iuvenum genua contingens sic adorabat :</p>	<p>English</p> <p>Now an old man appeared , gazing down on us from a summit at hand ; a goat-herd he was , as could be seen by the she-goats browsing round him . One of us asked him if he'd any milk or curds for sale . He shook his head several times before replying : 'How can you dream of food and drink , or anything else right now ? Don't you know where you are ? ' Then he gathered his goats , and made off into the distance . His words and his sudden flight filled us all with no little dread . We wondered what was wrong with the place , but there was no one the others could ask , till a second old man approached on the road , tall and bent with the years , hunched over his staff , wearily dragging his feet , and weeping copiously . Meeting with us he clasped the knees of all the young men in turn , wracked by tears .</p>	<p>English</p> <p>An old man appeared at the top of the hill with goats feeding around him . One of our people hailed him and asked whether he had any milk or fresh cheese for sale . He shook his head two or three times before answering : 'How can you think of food or drink or anything else of the sort ? Don't you know in what sort of a place you are camping ? ' He turned his back and went off with his goats . His question and the abrupt way he left us alarmed our people . They all began wondering what was wrong with the place . But there was nobody to enlighten them until another old man appeared , a tall , bent old man , dragging his feet wearily towards us and leaning heavily on a stick . When he reached the glade where we had halted , he fell down on his knees , his eyes streaming with tears , embraced our people one after the other and groaned</p>
---	--	---

Undergraduate, native English speaker, intermediate Latin. The student performed a three-text alignment of a passage from the *Metamorphoses* by Apuleius, comparing the original and two chosen English translations (A.S. Kline, 2016 and R. Graves, 1950), to discover differences in translation strategies in a very complex and rhetorically skilful literary work. The focus of the observation was the semantics of different words and expressions with potential ambiguity, but also the different ways the translators addressed the challenges of rendering Latin syntactical structures in a completely different linguistic system. The student then created a new English translation, discussing where it distanced itself from the original and which aspects were instead retained. The comparison between two very different authorial translations proved useful to understand different strategies of addressing relevant linguistic and semantic problems, but also to help the student to find their own “translator’s voice”.

26

Example 3: Homer, *Odyssey* 9.105-115. Ancient Greek-English.^[11]

<p>English</p> <p>Thence we sailed on , grieved at heart , and we came to the land of the Cyclopes , an overweening and lawless folk , who , trusting in the immortal gods , plant nothing with their hands nor plough ; but all these things spring up for them without sowing or ploughing , wheat , and barley , and vines , which bear the rich clusters of wine , and the rain of Zeus gives them increase . Neither assemblies nor council have they , nor appointed laws , but they dwell on the peaks of lofty mountains in hollow caves , and each one is lawgiver to his children and his wives , and they reckon nothing one of another .</p> <p>(20) 16% (103) 84% ENG - GRC</p>	<p>Ελληνικά</p> <p>ἐνθεν δὲ προτέρω πλέομεν ἀκαχήμενοι ἦτορ : Κυκλώπων δ' ἐς γαῖαν ὑπερφιάλων ἀθελίστων ἰκόμεθ' , οἳ ῥα θεοῖσι πεποιθότες ἀθανάτοιαν οὔτε φυτεύουσιν χερσὶν φυτὸν οὔτ' ἀρόωσιν , ἀλλὰ τὰ γ' ἄσπαρτα καὶ ἀνήροτα πάντα φύονται , 110πυροὶ καὶ κριθαὶ ἢ δ' ἄμπελοι , αἳ τε φέρουσιν οἶνον ἐριστάφυλον , καὶ σφιν Διὸς ὄμβρος ἀέξει . τοῖσιν δ' οὔτ' ἀγοραὶ βουλευφόροι οὔτε θέμιστες , ἀλλ' οἳ γ' ὑψηλῶν ὄρεων ναῖουσι κάρηνα ἐν σπέσσι γλαφυροῖσι , θεμιστεύει δὲ ἕκαστος 115παῖδων ἢ δ' ἀλόχων , οὐδ' ἀλλήλων ἀλέγουσιν .</p> <p>(66) 79% ENG - GRC (18) 21% GRC</p>	<p>English</p> <p>We sailed on , our morale sinking , And we came to the land of the Cyclopes , Lawless savages who leave everything Up to the gods . These people neither plow nor plant , But everything grows for them unsown : Wheat , barley , and vines that bear Clusters of grapes , watered by rain from Zeus . They have no assemblies or laws but live In high mountain caves , ruling their own Children and wives and ignoring each other .</p> <p>(66) 79% ENG - GRC (18) 21% GRC</p>
---	---	--

Undergraduate, English-Chinese speaker, upper-elementary Greek. The student aligned the original text of *Odyssey* 9.105-115 against two different English translations, in prose and poetry respectively (A.T. Murray, 1919 and S. Lombardo, 2000). The student used the correspondences in meaning and words across the two English renderings to individuate possible matching words in the Greek, specifically focusing on the potentially discrepant translation strategies between prose and verse. Overall, the prose translation revealed a slightly better alignment rate, but typical Greek stopwords, such as *δέ* or *τε*, were consistently not translated in both English versions.

27

Case 2: Students of a third language

The typical case for trilingual alignment was a student who was proficient in two languages, and wished to improve a third one. The student would perform a trilingual alignment (possibly preceded by a pairwise alignment between the two better-known languages), systematically comparing a lesser-known language against two translations. The student would use the knowledge of the first two languages to leverage the obstacles in understanding the third one, by recognizing common syntactical patterns, morphologies, or similarities in vocabulary.

28

Interestingly, this case was not limited to the study of historical languages: a collateral advantage of trilingual alignment was that students with proficiency in a historical language could use this opportunity to focus on another language that they knew less, often a second modern language that they were studying at the same time.

29

Example 1: Tacitus, *Germania* 1.1. English-German-Latin.^[12]

English

Germany is separated from the Galli, the Rhæti, and Pannonii, by the rivers Rhine and Danube; mountain ranges, or the fear which each feels for the other, divide it from the Sarmatæ and Daci. Elsewhere ocean girds it, embracing broad peninsulas and islands of unexplored extent, where certain tribes and kingdoms are newly known to us, revealed by war. The Rhine springs from a precipitous and inaccessible height of the Rhætian Alps, bends slightly westward, and mingles with the Northern Ocean. The Danube pours down from the gradual and gently rising slope of Mount Abnoba, and visits many nations, to force its way at last through six channels into the Pontus; a seventh mouth is lost in marshes.

(15)

(121) 89% ENG - DEU

Deutsch

Ganz Germanien wird von den Galliern und Rättern und von den Pannoniern durch die Flüsse Rhein und Donau, von den Sarmaten und Daziern durch wechselseitige Furcht oder Gebirge getrennt; das Übrige umgibt der Ozean, weite Buchten und unermessliche Inselräume umfassend, wobei man vor Kurzem von gewissen Völkern und Königen erfahren hat, die der Krieg ans Licht brachte. Der Rhein, entsprungen aus dem unerreichbaren und steilen Gipfel der rätischen Alpen, mischt sich, nachdem er sich mit einer mäßigen Beugung gen Westen gewandt hat, mit dem nördlichen Ozean. Die Donau, aus dem sanften und mild herausragenden Gipfel des Schwarzwaldes herausströmend, sucht mehrere Völker auf, bis sie sich über sechs Läufe ins Schwarze Meer ergießt: die siebte Mündung wird von Sümpfen verschlungen.

(125) 93% ENG - DEU

(10)

Latin

Germania omnis a Gallis Raetisque et Pannoniis Rheno et Danubio fluminibus, a Sarmatis Dacisque mutuo metu aut montibus separatur: cetera Oceanus ambit, latos sinus et insularum immensa spatia complectens, nuper cognitiss quibusdam gentibus ac regibus, quos bellum aperuit. Rhenus, Raeticarum Alpium inaccessio ac praecipiti vertice ortus, modico flexu in occidentem versus septentrionali Oceano miscetur. Danubius molli et clementer edito montis Abnoba iugo effusus plures populos adit, donec in Ponticum mare sex meatibus erumpat; septimum os paludibus hauritur.

(125) 93% ENG - DEU

(10)

Undergraduate, native English speaker, advanced Latin, basic German. The student performed a trilingual alignment of Tacitus' *Germania*, against an English and a German translation, and used his knowledge of Latin to improve his mastery in German. In his final report, he observed higher matching rate between the two inflected languages than with English: rather paradoxically, comparing German and Latin resulted easier than English and Latin. The English translation chosen was, on the other hand, approached with some criticism, because of its lengthy rendering of expressions that resulted equivalent in German and Latin.

30

Example 2: Loqman Fables. Arabic-French-Ancient Greek.[13]

العربية

أَسَدٌ مَرَّةً خَرَجَ عَلَى ثَوْرَيْنِ فَاجْتَمَعَا جَبِيماً وَكَانَا يَنْتَحَايِيهِ بِقُرْبَيْهِمَا وَلَا يُؤْتِكُمَا مِنَ الْخُجُولِ بَيْنَهُمَا فَتَقَوَّ بِأَحْدِهِمَا وَخَدَمَهُ وَبَعْدَهُ بَأَنَّ لَا يُعَارِضُهُمَا إِنْ نَحَلَى عَنْ صَاحِبِهِ فَلَمَّا اقْتَرَبَا اقْتَرَسَهُمَا جَبِيماً

هَذَا مَعْنَاهُ إِنْ عَرِيتَيْنِ إِذَا اتَّقَى رَأَى أَهْلَهُمَا فَإِنَّهُ لَا يُتَكَنُّ مِنْهُمَا عَدُوٌّ فَإِذَا اقْتَرَبَا هَلِكَا جَبِيماً

(49) 100% ARA - FRA

français

Un lion et deux taureaux.

Un lion attaqua un jour deux taureaux. Ceux-ci s'unirent ensemble, et, le visant de leurs cornes, le mirent hors d'état de pénétrer au milieu d'eux. Alors le lion, s'adressant à un des taureaux à part, le trompa par la promesse qu'il ne lui ferait rien, si celui-ci se séparait de son compagnon. Mais dès qu'ils furent séparés le lion les déchira tous deux.

Cela signifie qu'aucun ennemi ne peut se rendre maître de deux villes tant que l'avis de leurs habitants est le même; mais dès qu'elles sont séparées, elles périssent ensemble.

(115) 89% ARA - FRA

(14)

Ελληνική

Λέων και δύο ταύροι.

Λέων προσβαλών ταύρους δυοίν, ἐπειράτο τούτους καταβουήσασθαι, οἱ δὲ τὰ ἑαυτῶν κέρατα ἐπίσης αὐτῷ ἀντιπαράταξαν οὐκ εἶον τὸν λέοντα μέσον αὐτῶν παρελθεῖν. Ὁρῶν τοίνυν ἐκεῖνος, ὡς ἀδυνατῶς ἔχει πρὸς αὐτούς, κατεσφίσατο τοῦ ἐνόος, καὶ πρὸς αὐτὸν ἔφησεν, ὡς εἶγε τὸν σὸν ἑταῖρον προβιάσις μοι, ἀβλαβῆ σε διατηρήσω, καὶ τούτῳ τῷ τρόπῳ ἀμφοτέρους τοὺς ταύρους ἀνήρκεν.

Ἄ μῦθος δηλοῖ, ὡς ὅτι καὶ πόλεις καὶ ἄνθρωποι ἀλλήλοις ὁμοσούντες, οὐ συγχωροῦσιν αὐτῶν τοὺς ἐχθρούς περιγίνεσθαι. τῆς δ' ὁμοιοῦς περιφρονοῦντες, εὐχερῶς ἄρδην τοῖς ἐναντίοις ἀλίσκονται.

(115) 89% ARA - FRA

(14)

Graduate, advanced French, intermediate Ancient Greek, basic Arabic. The student performed a trilingual alignment of the famous Loqman fables, to pursue a research project that focused on gathering systematic evidence of the relation between Loqman and the Aesopic corpus, which the sage is often said to have translated [Heller and Stillmann 2012]: the student aligned the original Arabic and the corresponding text of the fables from the Aesopic corpus, using the French translation of Loqman as a pivot between the two languages. The high matching rate across the three versions, and the often verbatim correspondences between the Arabic version and the Aesopic fables, provide the first dataset of analytically collected evidence of the dependence between the Loqman and the Aesopic corpus.

31

Example 3: Sophocles, *Ajax* 646-692. Ancient Greek-Chinese-English.[14]

Ελληνική	中文	English
<p>Αἴας</p> <p>ἀπανθ' ὁ μακρὸς κἀναριθμητὸς χρόνος φύει τ' ἀδηλα καὶ φανέντα κρύπτεται : κούκ' ἔστ' ἄελλπον οὐδέν , ἀλλ' ἀλίσκεται χῶ δεινὸς ὄρκος χαί περισκελεῖς φρένες .</p> <p>κἀνὼ γάρ , ὃς τὰ δειν' ἐκαρτέρουν τότε , βαφῆ σιδήρος ὡς ἐθελώνθην στόμα πρὸς τῆσδε τῆς γυναίκας : οἰκτίρω δέ νιν χῆραν παρ' ἐχθροῖς παῖδά τ' ὄρφανὸν λιπεῖν .</p> <p>ἀλλ' εἶμι πρὸς τε λουτρά καὶ παρακτίους λειμώνας , ὡς ἂν λύμαθ' ἀγνίσας ἐμὰ μῆνιν βαρεῖαν ἐξαλύξωμαι θεάς : μολῶν τε χῶρον ἐνθ' ἂν ἀσιτῆρῆ κίχῳ , κρύψω τόδ' ἔγχος τούμῳν , ἐχθιστον βελῶν , γαίας ὀρούσας ἐνθα μὴ τις ὄψεται : ἀλλ' αὐτὸ νύξ 'Αΐδης τε σφζόντων κάτω .</p>	<p>埃阿斯</p> <p>漫长而数不尽的岁月揭示了所有混沌 ， 然后在它显现后，再次隐藏了它。 没有什么是无法预料的， 庄严的誓言与最倔强的心灵都被征服了。</p> <p>同样的，我曾经历尽艰险，如淬火的 钢铁般坚硬， 如今我的刃也被这个女人给软化了。 我于心不忍让她成为敌人堆里的寡妇 ， 或让我的孩子成为孤儿。</p> <p>但是我走向沐浴之处和海岸边的草地 ， 为了逃脱女神的暴怒， 我洗净了我的污秽。</p>	<p>Ajax</p> <p>All things the long and countless years first draw from darkness , and then bury from light ; and there is nothing which man should not expect : the dread power of oath is conquered , as is unyielding will .</p> <p>For even I , who used to be so tremendously strong – yes , like tempered iron–felt my tongue ' s sharp edge emasculated by this woman ' s words , and I feel the pity of leaving her a widow and the boy an orphan among my enemies .</p> <p>But I will go to the bathing-place and the meadows by the shore so that by purging my defilements I may escape the heavy anger of the goddesses . Then I will find some isolated spot , and bury this sword of mine , most hateful weapon , digging down in the earth where none can see . Let Night and Hades keep it underground !</p>

Graduate, Chinese native speaker, English as second language, intermediate Ancient Greek. The student performed a trilingual alignment of Sophocles' *Ajax*, aligning the Greek text, the English translation by Sir R.C. Jebb (1896), and a Chinese version based on the English. While the Chinese and the English obviously displayed a higher matching rate, the student was able to use the knowledge of both languages to individuate correspondences with the Greek, and also to establish where there was no available match for Greek words or concepts.

32

Case 3: No prior knowledge of the source language

The third case involved students, enrolled in courses on culture or literature, who had no prior knowledge of the source language and were also not studying it. The students were given short sections of a source text and corresponding translations in various languages, and performed several tasks aimed at gradually building a critical approach to the original:

33

1. Individuate single words in the source text, such as “holy”, “honor”, or “love”, and align them with the corresponding words in a translation in a chosen language. The goal of this exercise was to develop an understanding of the depth of crucial cultural concepts, by assessing the many different ways in which the same word in the source text could be translated.
2. Compare how specific expressions were rendered in two translations, and then align them against the source text.^[15] The students focused on different ways of conveying morphosyntactic constructs across languages, critically approaching different translation strategies.
3. Complete bilingual or trilingual alignment of a short passage, with intensive and systematic analysis of the linguistic and expressive content. The purpose of this exercise was to let students assess the linguistic differences between the languages in question, and partly to demonstrate to them that they could, in fact, approach a text in a language that they did not know at all, by means of translation alignment.

Interestingly, we observed that students who had mastery of more than one language almost immediately learned to use trilingual alignment to leverage the difficulties in dealing with a language that they had never seen before. This strategy proved effective in various cases: many students completed the first two assignments using both English and a third modern language (often Spanish or French); non-native English speakers used their native language to better understand complex words and expressions; students who mastered another classical language tested if they could use it to better engage with phenomena like inflected forms, participial uses, and subordination. Overall, trilingual alignment proved to be an effective strategy to engage with the source text, as already assessed in Case 2. By the end of the course, students were able to align relatively long passages against one or more translations.

Example 1: Herodotus, *Histories* 1.35. Ancient Greek-English-Latin.^[16]

Ελληνική	English	Latin
<p>Ἐχοντας δὲ οἱ ἐν χερσὶ τοῦ παιδὸς τὸν γάμον ἀπικνέεται ἐς τὰς Σάρδεις ἀνὴρ συμφορῇ ἐχόμενος καὶ οὐ καθαρὸς χεῖρας, ἐὼν Φρύξ μὲν γενεῇ, γένεος δὲ τοῦ βασιλίου. Παρθένων δὲ οὗτος ἐς τὰ Κροίσου οἰκία κατὰ νόμου τοὺς ἐπιχωρίους καθαρσίω ἐδέετο κυρῆσαι, Κροῖος δὲ μιν ἐκώθηρε. Ἔστι δὲ παραπλήσιή ἡ καθαρσις τοῖσι Λυδοῖσι καὶ τοῖσι Ἕλλησι. Ἐπειτέ δὲ τὰ νομιζόμενα ἐποίησε ὁ Κροῖος, ἐπυθάνετο ὁκόθεν τε καὶ τίς εἴη, λέγων τῷδε· Ὀρθρωπε, τίς τε εὖν καὶ κόθεν τῆς Φρυγίης ἦκων ἐπιστάσις μοι ἐγένεο; Τίνα τε ἀνδρῶν ἢ γυναικῶν ἐφόνευσας; Ὁ δὲ ἀμείβετο· Ὡ βασιλεῦ, Γορδίω μὲν τοῦ Μίδεω εἰμι παῖς, ὀνομαζόμενος δὲ Ἄδραστος, φονεῦσας δὲ ἀδελφεὸν ἐμειωτοῦ ἀέκωμ πάρεμι ἐξεληλαμένους τε ὑπὸ τοῦ πατρὸς καὶ ἐστερημένους πάντων. Κροῖος δὲ μιν ἀμείβετο τοῖοδε· Ἄνδρῶν τε φίλων τυγχάνεις ἔκγονος ἐὼν καὶ ἐλήλυθας ἐς φίλους, ἐνθα ἀμικχανήσεις χρηματός οὐδενός μὲν ἐν ἡμετέρου συμφορῆν τε ταύτην ἢ κούφοτά φέρων κεραβανέας πλεῖστον.</p>	<p>Then while he was engaged about the marriage of his son, there came to Sardis a man under a misfortune and with hands not clean, a Phrygian by birth and of the royal house. This man came to the house of Croesus, and according to the customs which prevail in that land made request that he might have cleansing; and Croesus gave him cleansing: now the manner of cleansing among the Lydians is the same almost as that which the Hellenes use. So when Croesus had done that which was customary, he asked of him whence he came and who he was, saying as follows: "Man, who art thou, and from what region of Phrygia didst thou come to sit upon my hearth? And whom of men or women didst thou slay?" And he replied: "O king, I am the son of Gordias, the son of Midas, and I am called Adrastus; and I slew my own brother against my will, and therefore am I deprived of all that I had." And Croesus answered thus: "Thou art, as it chances, the offshoot of men who are our friends and thou hast come to friends, among whom thou shalt want of nothing so long as thou shalt remain in our land: and thou wilt find it most for thy profit to bear this misfortune as lightly as may be." So he had his abode with Croesus.</p>	<p>Interim, dum nuptias filii parat Croesus, advenit Sardes vir calamitate obstrictus, cui manus non erant purae, natione Phryx, regio de genere. Qui quum ad aedes Croesi venisset, orassetque ut ritu patrio expiari sibi contingeret, expiavit eum Croesus: est autem modus expiandi apud Lydos similis ei, quo Graeci utuntur. Peractis legitimis, percontatur Croesus hominem, unde veniret, quisve esset, his usus verbis: "Quis tu es, homo, et quonam e Phrygiae loco veniens, ad meos lares supplex te recepisti? et quem virum, quamve foeminam occidisti?" Cui ille respondit: "O rex, Gordiae sum filius, Midae nepos; est autem nomen mihi Adrasto. Fratrem meum occidi invitus; unde ejectus a patre adsum, et rebus omnibus destitutus." Tum Croesus: "Ex viris amicis, inquit, oriundus es, et ad amicos venisti, ubi nullius rei indigebis quoad apud nos manseris. Istam autem calamitatem quam poteris levissime ferendo, plurimum lucri feceris, Ita ille in Croesi aedibus vitam agebat."</p>

Undergraduate, native English speaker, advanced Ancient Greek, no Latin. The student used trilingual alignment to verify whether the knowledge of Greek, with English as a bridge language, could serve to leverage the unfamiliarity with a text in Latin. The passage chosen, from Herodotus' *Histories*, was aligned in the original Greek, against the English translation by A.D. Godley (1920), and the Latin version by J. Schweighaeuser (1824). The experiment proved successful: the student was able to recognize cognate words and match them with the corresponding Greek, and to use the knowledge of inflected languages to compare parts of speech between Latin and Greek (for example, to identify the accusative *fratre-m* based on the similar ending with the Ionic Greek ἀδελφέ-ν).

34

Example 2: Euripides, Bacchae 395-399. Ancient Greek-Chinese-English.^[17]

Ελληνική	English	中文
<p>τὸ σοφὸν δ' οὐ σοφία τὸ τε μὴ θνητὰ φρονεῖν. βραχύς αἰὼν: ἐπὶ τούτῳ δὲ τις ἂν μεγάλα διώκωσ τὰ παρόντ' οὐχὶ φέροι.</p>	<p>But being clever isn't wisdom. And thinking deeply about things isn't suitable for mortal men. Our life is brief - that's why the man who chases greatness fails to grasp what's near at hand.</p>	<p>但显得聪明不是智慧。 且凡人不适合 对事物追究过深 人生稍纵即逝 - 因此追求伟大之人 会让唾手可得之物溜走。</p>
(8) 31% GRC	(18) 69% GRC - ENG	(25) 68% GRC - ENG
	(12) 32% ENG	(25) 68% GRC - ENG
		(12) 32% ENG

Undergraduate, Chinese native speaker, English as second language, no Ancient Greek. The student created a trilingual alignment of Euripides' Bacchae, comparing the English translation by I. Johnson (<http://jelks.nu/libri/classics/bacchae.html>) and a revised Chinese version based on the English (as there were no available Chinese translations of the original Greek). As a Chinese native speaker, the student used Chinese as a bridge to establish accurate correspondences between the English and the Ancient Greek words that had never been seen before. The result was an investigation into the meaning of Ancient Greek concepts through a meaningful comparison with similar Chinese terms, often with an emphasis on imperfect grammatical and semantic correspondences between the Greek and its translation (such as τὸ σοφόν, "cleverness", translated as "being clever").

35

Observations

At the end of the course all the students, no matter what level of mastery they had of the languages, developed an acute sense of how limited their understanding of the source text was from the translation, as much as the critical recognition that they could convey in-depth knowledge of critical cultural concepts expressed in the source language, by contrasting them with their modern language translation and investigating into their various meanings [Crane 2019]. The recognition of the limited capabilities of translation to efficiently convey concepts in Classical languages, while at first caused disappointment and criticism, convinced the students of the need to look more in depth into the original language to understand its inherent characteristics and the way it expressed ideas [Palladino 2020].

36

Students of the source language developed a tangible sense for the fluidity of translation by evaluating different

37

strategies employed by professional scholars to approach complex phenomena; they systematically approached complex morphosyntactic constructs and were compelled to discuss their function and significance, while assessing the necessary imperfect character of any translation of them. In addition, students who were experimenting with learning a new modern language could approach unknown expressions and vocabulary by using their skills in other languages as a bridge to convey similar constructs.

One of the most relevant outcomes of this study was that students with no prior knowledge of the source language could start learning it by directly reading original literary texts. This approach could be revolutionary in the field of slow reading and language learning, and it also promises non-trivial consequences for passive users: readers with no knowledge of a source language could easily use already available alignments to perform a dynamic reading of the original aligned with its translation, gaining a basic understanding of vocabulary and syntax in the original.

38

Limitations and Future Work

We hope to have shown the pedagogical potential of translation alignment. However, we are also aware that significant implementation needs to be pursued to improve some underdeveloped aspects and to allow a more systematic use of a translation alignment interface in teaching practices.

39

We recognize that at the moment the tool is open to every category of user, and the aligned pairs are partly the result of unsupervised work, which may affect the quality and consistency of the information collected in the database. The lack of a suitable evaluation/correction workflow not only makes the dataset not accurate enough for machine translation, but also prevents organic teacher supervision. Therefore, the integration of a voting system for the evaluation of the accuracy by teachers or expert users is one of the most needed implementations. A related problem is the current impossibility of working on collaborative projects, where multiple users can work together on the same corpus.

40

Secondly, not all translation pairs are equally useful as training data. We should expect alignment to produce conflicting, mutually exclusive, results: after all, translation alignment is the result of interpretation, and not all cases are easily classifiable as “right” or “wrong”. In some cases, typically literary texts, it is extremely difficult to establish perfect or even partial correspondences between expressions and concepts in potentially very different languages: we have found that, in lack of specific guidelines or gold standards for each single language pair, users tend to create their own set of rules to keep a consistent alignment strategy, but these rules tend to be very different according to the particular purpose of the alignment. Therefore, we need a system to filter translation pairs: an option would be to allow users to classify different kinds of translation pairs, for example distinguishing categories such as “perfect/complete” or “partial/incomplete”, and “literal” or “free/poetical”.

41

Finally, we recognize that immediate access to part-of-speech tagging and available word pairs while performing the alignment would enormously impact on the work of non-specialized users and students, and obviously on the improvement of analytical passive reading. This very desirable additional feature is also included in our future implementation.

42

Notes

[1] In digital environments, annotation is the operation of attaching additional information to a resource or a specific part of it: an annotation can be unstructured, such as a comment, or structured, such as a machine-actionable citation identifying a text [Romanello and Pasin 2013].

[2] A text and its translation or translations placed alongside are defined as parallel texts. Large collections of parallel texts are called parallel corpora, such as the European Parliament Proceedings (<http://www.statmt.org/europarl/>) or the Bible Corpus [Christodouloupoulos and Steedman 2015]. Large research infrastructures like CLARIN (<https://www.clarin.eu/resource-families/parallel-corpora>) host big collections of parallel corpora in multiple languages as a resource for automatic translation and linguistic research. The use of parallel corpora in translation management systems is increasingly popular: commercial tools like SDL Trados (<https://www.sdltrados.com/>), MemoQ (<https://www.memoq.com/>), or Across (<https://www.across.net/>) allow professional translators to manually align sentences while performing their work, at the same time keeping consistency in the translation.

[3] For current issues in using Glosbe for statistical machine translation, see Berti (2015).

[4] Since the short vowels in Arabic and Persian are not written, the model used here needs to be improved to increase the readability of the transliterated text.

[5] GIZA++ (<http://statmt.org/amos/giza/GIZA++.html>) is a popular program, created as an implementation of the famous IBM models 1-5, which provides automatic alignments of words or phrases to implement statistical machine translation.

[6] See, for instance, Example 1 of Case 1, where alignment was part of a complete digital project.

[7] In 2018, Furman students presented their projects at two undergraduate conferences, at UNC Greensboro and University of South Carolina respectively.

[8] <http://ugarit.ialigner.com/text.php?id=2154>

[9] [https://github.com/ChiaraPalladino/TuftsDCC/wiki/Jiyoung-Song,-Diotima-on-Love-\(Plato,-%22Symposium%22\)](https://github.com/ChiaraPalladino/TuftsDCC/wiki/Jiyoung-Song,-Diotima-on-Love-(Plato,-%22Symposium%22))

[10] <http://ugarit.ialigner.com/text.php?id=25919>

[11] <http://ugarit.ialigner.com/text.php?id=26615>

[12] <http://ugarit.ialigner.com/text.php?id=26669>

[13] <http://ugarit.ialigner.com/text.php?id=24421>

[14] <http://ugarit.ialigner.com/text.php?id=25279>

[15] A strategy fully developed in Case 1, *above*.

[16] <http://ugarit.ialigner.com/text.php?id=25197>

[17] <http://ugarit.ialigner.com/text.php?id=2134>

Works Cited

- Babeu 2019** Babeu, A. "The Perseus Catalog: of FRBR, Finding Aids, Linked Data, and Open Greek and Latin." In Berti, M., *Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution*. De Gruyter, Berlin-Boston (2019): 53-72.
- Berti and Yousef 2015** Berti, M. and Yousef, T. "The Digital Fragmenta Historicorum Graecorum and the Ancient Greek-Latin Dynamic Lexicon." In Passarotti, M., Mambrini, F., and Sporleder, C. (eds), *Proceedings of the Workshop on Corpus-Based Research in the Humanities (CRH)*, 10 December 2015 Warsaw, Poland. Institute of Computer Science: Polish Academy of Sciences, Warsaw (2015): 117-123.
- Brown et al. 1990** Brown, P. F. et al. "A Statistical Approach to Machine Translation." *Computational Linguistics* 16.2 (1990): 79-85.
- Christodouloupoulos and Steedman 2015** Christodouloupoulos, C. and Steedman, M. "A Massively Parallel Corpus: the Bible in 100 Languages." *Language Resources and Evaluation*. Vol. 49.2 (2015): 375-295.
- Crane 2019** Crane, G. R. "Confronting Complexity of Babel in a Global and Digital Age. How can you work with a language that you do not know?" DH2019: Digital Humanities Conference, University of Utrecht, July 9-12. Book of Abstracts. <https://dev.clariah.nl/files/dh2019/boa/0611.html>.
- Dagan et al. 1999** Dagan, I. et al. "Robust Bilingual Word Alignment for Machine Aided Translation." In Armstrong, S. et al. (eds), *Natural Language Processing Using Very Large Corpora*. Springer Netherlands, Dordrecht (1999): 209-224.
- Foradi 2019** Foradi, M. "Confronting Complexity of Babel in a Global and Digital Age. What can you produce and what can you learn when aligning a translation to a language that you have not studied?" DH2019: Digital Humanities Conference, University of Utrecht, July 9-12. Book of Abstracts. <https://dev.clariah.nl/files/dh2019/boa/0611.html>
- Foradi et al. 2017** Foradi, M., Yousef, T., and Palladino, C. "*IAligner e o Editor Ugarit Em Projetos de Ensino - Aprendizagem e Alinhamento de Tradução*." Seminar presented at "*Do Texto Clássico ao Cibertexto: 50 anos do Curso de Grego Antigo de Araraquara*." UNESP Araraquara and Mackenzie University São Paulo (2017).
- Graça et al. 2009** Graça, J. et al. "Building a Golden Collection of Parallel Multi-Language Word Alignment." LREC (2008).

- Heller and Stillmann 2012** Heller, B. and Stillmann, N. A. "Luḡmān." In Bearman, P. et al. (eds) *Encyclopedia of Islam, Second Edition*. Brill Online Reference Works (2012). Accessed May 2019. http://dx.doi.org/10.1163/1573-3912_islam_COM_0586.
- Kay and Röscheisen 1993** Kay, M. and Röscheisen, M. "Text-translation alignment." *Computational Linguistics* 19.1 (1993): 121-142.
- Kumar et al. 2007** Kumar, S., Och, F., and Macherey, W. "Improving Word Alignment with Bridge Languages." In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. Association for Computational Linguistics, Prague (2007): 42-50.
- Melamed 1998** Melamed, I.D. "Manual Annotation of Translational Equivalence: The Blinker Project." *IRCS Technical Reports Series* (1998). https://repository.upenn.edu/ircs_reports/54.
- Palladino 2020** Palladino, C. "Reading Texts in Digital Environments: Applications of Translation Alignment for Classical Language Learning." *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/>
- Romanello and Pasin 2013** Romanello, M. and Pasin, M. "Citations and Annotations in Classics: Old Problems and New Perspectives." *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: Metadata, Vocabularies and Techniques in the Digital Humanities*, ACM, New York (2013).
- Véronis 2000** Véronis, J. *Parallel Text Processing: Alignment and Use of Translation Corpora*. Springer Netherlands, Dordrecht-Boston-London (2000).
- Wolf 2018** Wolf, M. *Reader, Come Home. The Reading Brain in the Digital World*. Harper, New York (2018).
- Yousef 2015** Yousef, T. *Word Alignment and Named-Entity Recognition Applied to Greek Text-Reuses*. Master Thesis, University of Leipzig (2015).
- Yousef 2019** Yousef, T. "Ugarit: Translation Alignment Visualization." In *LEVIA'19: Leipzig Symposium on Visualization in Applications 2019*. Leipzig.
- Yousef and Jänicke 2020** Yousef, T., and Jänicke, S. "A Survey of Text Alignment Visualization." *IEEE Transactions on Visualization and Computer Graphics* PP (October): 1–1. <https://doi.org/10.1109/TVCG.2020.3028975>.