

That branch of the Lake of Como...: Developing a New Resource for the Analysis of *I Promessi Sposi* and its Historical Translations

Rachele Sprugnoli^{1,*†}, Marco Sartor^{1,†}

¹Università di Parma, Viale D'Azeglio, 85, 43125 Parma, Italy

Abstract

This paper presents a directional parallel corpus of the Ventisettana, that is the version of *I Promessi Sposi* published by Manzoni in 1827, aligned at sentence level with the anonymous English translation published in London in 1834 by Richard Bentley. After describing the procedure followed for creating the resource and analyzing the results of the manual alignment, the corpus is used as a gold standard to evaluate Bertalign automatic aligner. This new linguistic resource can benefit the research community, in particular in the fields of the history of literature and translation studies, and be useful for developing new automatic tools specific for handling the peculiarities of historical literary texts.

Keywords

parallel corpus, sentence alignment, translation, digital humanities

1. Introduction

Critics have established how the *Promessi sposi* immediately enjoyed a wide resonance in Europe and in US, although the success of the work outside Italy has not always been accompanied by an effective understanding of the author's thought.¹ For this reason, the development of new linguistic resources based on the first historical translations of the novel assumes particular importance. These resources will benefit the research of historians of the Italian language, but also the development of new automatic tools suitable for processing historical literary texts. Last but not least, they can be used for educational purposes, both in secondary school, for the study of Manzoni's texts and their circulation beyond national borders, and at university level, in the field of translation studies. In particular, in this contribution we present a parallel corpus of the so-called Ventisettana, that is the version of the novel published by Manzoni in 1827, aligned with the anonymous English translation published in London in

1834 by Richard Bentley. An analysis of these two texts and of other English translations of the 19th century is provided by Intonti and Mallardi [5]: the volume is accompanied by examples of alignments to show specific cases both at the level of sentences, such as cuts and additions, and at the level of words, such as the rendition of figurative expressions and proverbs. The creation of a more extensive resource, such as the one presented here, aims to test the feasibility of a procedure to be applied in the future also to other historical translations so to offer the possibility of extending the range of linguistic analysis. Furthermore, our parallel corpus is a gold standard for evaluating fully automatic algorithms in a complex setting due to the peculiarities of historical texts and historical translations. Indeed, the complexity is due both to the characteristics of Manzoni's novel (rich, among other things, in irony, dialectal expressions, dialogues and monologues) and to the fact that during the 19th century translations did not aim to guarantee the greatest possible fidelity towards the source text, but rather to bend it in the light of the historical-cultural context in which they were implemented [6]. This approach to translation causes the original text to be revised and changed through additions and omissions of even entire chapters, making it a challenge to automate the alignment process.

2. Related Work

A parallel corpus is made of a set of texts in a given source language aligned with their translations in one or more target languages. The alignment, that is the identification of corresponding text units in parallel texts, can be performed at paragraph, sentence or word level. When

CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 – Dec 02, 2023, Venice, Italy

*Corresponding author.

[†]This paper is the result of the collaboration between the two authors. For the specific concerns of the Italian academic attribution system: Rachele Sprugnoli is responsible for Sections 2, 3.2, and 4; Marco Sartor is responsible for Section 3.1. Sections 1 and 5 were collaboratively written by Rachele Sprugnoli and Marco Sartor.

✉ rachele.sprugnoli@unipr.it (R. Sprugnoli);

marco.sartor@unipr.it (M. Sartor)

ORCID 0000-0001-6861-5595 (R. Sprugnoli); 0000-0002-1176-2735

(M. Sartor)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

¹A large number of blunders and mistakes made in the translations has been reported in [1] and [2]. For Manzoni's popularity outside Italy, see [3] and the references listed on [4].

the translation direction is known (i.e. when the source and target languages are clearly stated) and when the translation is direct (i.e. not mediated by an intermediary language), the parallel corpus is defined as directional [7].

The development of large parallel corpora, both bilingual and multilingual, took off in the 90s of the last century but their growth in terms of number of texts and languages covered is more recent thanks to initiative such as the OPUS project [8] and those promoted by the European Commission [9]. The great attention given to this type of corpora is due to the fact that parallel corpora are useful to gain insights into interlinguistic phenomena; at the same time they are a rich source of materials for language teaching, translation studies, lexicography, and a fundamental resource for terminology extraction and machine translation systems.

Since manual alignment is a particularly time-consuming process, various automatic techniques have been proposed over the years [10]. Specifically, with regard to sentence-level alignment, early approaches are based on sentence length in terms of number of words or characters. The idea behind this method is that long sentences in the source text are translated with long sentences, while short sentences are translated using short sentences [11, 12]. Lexical matching methods using bilingual dictionaries (such as in the hunalign system [13]) or specific tokens (such as dates, proper nouns, punctuation) as anchors for the alignment [14] are also worth mentioning. On the other hand, MT-based approaches require the source text to be automatically translated into the target language and use a similarity score (e.g. the BLEU metric) to align the machine translation output with the target text sentences; an example of this kind of method is given by Bleualign [15]. The most recent systems, however, are those based on multilingual sentence embeddings, such as Vecalign [16], or sentence-transformers, as Bertalign [17]. Such approaches have been tested on literary texts obtaining good performances [18].²

In this paper we present a manually created bilingual (IT-EN) directional parallel corpus of historical literary texts together with the evaluation of automatic sentence alignment methods. Dealing with texts written in not contemporary languages and of a literary genre is particularly interesting and not so widespread; suffice it to say that the CLARIN infrastructure gives access to 87 parallel corpora:³ out of these, only 5 include texts in Italian, but none contain works by Manzoni or historical literary translations.

²Results obtained on literary and non-literary texts using various methods, including the Vecalign and Bertalign systems, are reported in <https://github.com/bfsujason/aligner-eval>.

³<https://www.clarin.eu/resource-families/parallel-corpora>.

3. Dataset Creation and Analysis

3.1. Creation

The digital text of the Ventisettana was provided by the Italian project (PRIN 2017) *ManzoniOnline2: new documents, translations and tradition* [19],⁴ whereas the text of the 1834 English translation was downloaded from the Gutenberg project website as UTF-8 text file.⁵ Both texts have been divided into chapters; for each of them the sentence-level alignment was completed semi-automatically, with manual correction of the output of the aligner. In the initial phase of our work we tested various tools which include graphical user interfaces for editing the automatic alignment, such as TAligner 3.0 [20], LF Aligner⁶ and InterText. More specifically, as stated in [21], extensive trials were conducted with LF Aligner, before the final choice fell on InterText because of the intuitiveness of its interface and the possibility of exporting in various formats [22].

Each chapter was loaded onto InterText in a separate file with one sentence per line. Sentence splitting was done manually: we tried various sentence splitting models but always obtaining low performances due to the peculiarity of the novel's punctuation and to an unconventional use of capital letters. Among the Universal Dependencies (UD) 2.10 models available in UDPipe [23], the best result was obtained with VIT with an accuracy of 39%. A better, but far from perfect, accuracy (64%) was registered with Stanza [24]. Overall, it can be remarked that automatic sentence splitting fails especially (but not exclusively) with punctuation marks that are no longer in use or with traditional punctuation marks employed in unusual contexts compared to today's custom. In particular, the use of hyphens – short and long – with different functions is very frequent in the 1834 English translation. Normally, the latter separate one sentence from the other, mostly marking the end of a direct speech,⁷ while the former convey a character's inner thoughts, include an aside, render a hesitation in direct speech or mark a pause of medium intensity without giving rise to a new sentence.⁸ Automatic splitting also displays glitches when dealing with inverted commas marking the start of a direct speech, the three suspension dots, and

⁴<https://www.alessandromanzoni.org/>

⁵<https://www.gutenberg.org/ebooks/35155>.

⁶<https://sourceforge.net/projects/aligner/>.

⁷For instance: "But, fair sirs, you are too just, too reasonable—" "But," interrupted the other comrade... (from chapter 1).

⁸Here are some examples, all taken from chapter 1: "for if you do, ehem!—you understand—the consequences would be the same as if you performed the marriage ceremony"; "the poor curate neither meddles nor makes—they settle their affairs amongst themselves, and then—then, they come to us, as if to redeem a pledge; and we—we are the servants of the public"; "but he will require reasons—and what can I say to him"; "... and he arose, continuing—"No! I'll take nothing, nothing?".

Table 1

Examples of different types of alignment.

	IT	EN
2:1	I due sposi rimasti promessi si trovarono in faccia Agnese, che arrivava trambasciata e affannosa. «Ah siete qui!» diss'ella traendo la parola a stento.	The two lovers (still lovers) stood before Agnes, who, alarmed and grieved, said, "Ah! you are here!"
1:2	«Come è andata? che cos'è la campana?»	How has it gone? Why did the bell ring?"
1:0	Mi par d'aver inteso...	
1:1	«A casa, a casa,» diceva Renzo, «prima che venga la gente.»	"Home, home!" said Renzo, "before the people gather."

155	★ ▶ I due sposi rimasti promessi si trovarono in faccia Agnese, che arrivava trambasciata e affannosa. ▶ «Ah siete qui!» diss'ella traendo la parola a stento.	▶ The two lovers (still lovers) stood before Agnes, who, alarmed and grieved, said, "Ah! you are here!"	✓
156	★ ▶ «Come è andata? che cos'è la campana?»	▶ How has it gone? ▶ Why did the bell ring?"	✓
157	★ ▶ Mi par d'aver inteso...		✓
158	★ ▶ «A casa, a casa,» diceva Renzo, «prima che venga la gente.»	▶ "Home, home!" said Renzo, "before the people gather."	✓

Figure 1: Alignments as displayed in the InterText interface: sentences without a 1:1 alignment are highlighted in yellow.

exclamation or question marks followed by a lower-case letter (which do not start a new sentence but denote a single flow of text). At the end of the manual sentence splitting procedure, we obtained 8,718 sentences for the Ventisettana and 7,484 sentences for the English translation.

In the following phase, we manually corrected the automatic alignment made by hunalign system integrated in InterText. On average, 3 hours of work were required for validating each chapter. Texts were then exported in three files: each chapter was saved as two independent XML files (one for the Italian text and one for the English translation) and their alignment was exported as a separate XML file containing pointers to the individual sentences of the two texts.

3.2. Analysis

The alignments produced can be categorized into the following different types:

- 1:1, i.e. one sentence is translated by one sentence. It should be noted that such correspondence is not necessarily a symptom of total fidelity, or rather of a linear (or even literal) translation of the subphrasal units. While respecting the boundaries of the sentence, in fact, there could be phenomena of expansion or synthesis. For example, in chapter VIII, a long sentence – with a simile used to indicate how the *Bravi* (hired assassins) were gathered in a courtyard by their leader emphasizing their animal nature – is strongly synthesized by removing the rhetorical figure altogether.
 - Ventisettana: *Come il cane che scorta un gregge di porci corre or qua or là a quei che*

*si sbandano, ne addenta uno per un'orecchia e lo tira in ischiera, ne spinge un altro col muso, abbaia ad un altro che esce di fila in quel momento, così il pellegrino acciuffa uno di coloro che già toccava la soglia e lo strappa indietro, caccia indietro col bordone uno e un altro che v'eran già presso, grida agli altri che scorrazzano senza saper dove, tanto che li raccolzò tutti nel mezzo del cortiletto.*⁹

– 1834 English translation: *He succeeded, however, in assembling them in the middle of the court-yard.*

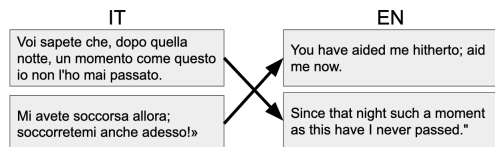
- 1:0 and 0:1, i.e. a sentence in the Ventisettana or in the translation lacks a parallel in the other text, following an omission (type 1:0) or an addition by the translator (type 0:1). Omissions are part of a wider trend in the historical translations of Manzoni's novel to significantly cut sentences that were considered not essential for understanding the text. This is aimed at giving the translation a drier and more pragmatic tone than the original, in line with the prevailing fashions in the literary context of reception; such approach is consistent with the so-called domestication strategy of translations [25].

⁹English literal translation: *Like the dog that escorts a herd of pigs, he runs here and there among those who are straying, he bites one by the ear and puts him in line, he pushes another with his muzzle, he barks at another who leaves the line at that moment, so the pilgrim grabs one of those who were already on the doorstep and snatches him back, he drives one and another who was nearby back with his stick, he shouts to the others who are running around without knowing where, so much so that he gathered them all in the middle of the little courtyard.*

Table 2

Number of alignments per type.

TYPES	1-1	1-0	2-1	1-2	3-1	1-3	OTHER
NUMBER OF ALIGNMENTS	5,068	1,077	693	597	106	84	108

**Figure 2:** Example of cross-order alignment.

- 1:N and N:1, i.e. the translator has split or merged the original sentences. When one Italian sentence is split into two or more sentences the alignment is 1:N. When, on the contrary, two or more Italian sentences are merged in a single sentence in the translation the alignment is N:1.

Table 1 provides examples, taken from chapter VIII, of the aforementioned types, while Figure 1 shows how the same alignments are displayed in InterText interface. In addition, Table 2 presents the number of alignments per type. The vast majority of alignments are 1:1 (66%), but there are also several omissions in the translation (1:0, 14%), followed by cases of 2:1 merging (9%) and 1:2 splitting (8%). Under the “Other” category we collect the types having a number of occurrences less than 1% (i.e. 0-1, 4-1, 1-4, 5-1, 6-1, 3-2, 1-5). It is important to notice that our resource includes few cases of cross-order alignments in which the translator has changed the order of the sentences in the translation so that, to create the alignment, it is necessary to move sentences out of their original position (which is possible with InterText). Cross-order alignments fall into the types described above: for example, Figure 2 shows a cross-order alignment, taken from chapter XXXVI, which generates a 1:1 match between the source and the target sentences.

4. Testing Automatic Alignment Methods

The parallel corpus described in the previous section has been used as gold standard for testing the performances of Bertalign, an automatic aligner that uses LaBSE (language-agnostic BERT sentence embeddings, [26]) for building cross-lingual embeddings of source and target sentences.¹⁰ As reported by Liu and Zhu [17], Bertalign is designed with the aim of dealing with non-1-

¹⁰<https://github.com/bfsujason/bertalign>

to-1 sentence pairs that are quite common in literary texts. The comparative evaluation carried out on literary texts considering the English-Chinese translation pair showed that Bertalign is able to outperform other (length-based, dictionary-based, MT-based and embedding-based) aligners.

We configured Bertalign with the following options:

- maximum alignment types (`max_align`): 6
- k nearest target neighbors of each source sentence (`top_k`): 3
- search window (`win`): 5
- similarity score for 1:0 and 0:1 alignments (`skip`): 0
- modified cosine similarity as proposed in [17] (`margin`): True
- length difference between source and target sentences (`len_penalty`): False
- sentence splitting (`is_split`): True

With respect to the default configuration, we increased the maximum alignment length (i.e. the `max_align` option) from 5 to 6 because our corpus has many complex alignments, that is various types of 1:N and N:1 alignments. We also set a larger value for the similarity score (i.e. the `skip` option) because our corpus contains many omissions and insertions. Given that we have several cases of expansion or synthesis even in 1:1 alignments, the `len_penalty` parameter is set to False: in this way the length difference between source and target sentences is not taken into consideration when calculating the similarity between sentence pairs. On the contrary, the `is_split` option is set to True because our corpus was already split into sentences.

Table 3 reports the results of our evaluation using both Bertalign (with the default configuration, *Bertalign_d*, and with our custom options, *Bertalign_c*) and the Galechurch length-based algorithm. The superiority of the embedding-based approach over the length-based one is evident: the former outperform the latter by 5 F1 points. The custom configuration further improves Bertalign’s performance in terms of both precision and recall. However, the results are slightly lower than those recorded on the English-Chinese pair: indeed, for the MAC corpus of literary texts a precision of 0.906, a recall of 0.912 and an F1 of 0.909 are reported.¹¹

Figure 3 displays F1 performance across the chapters of the novel. The variation between individual chapters

¹¹<https://github.com/bfsujason/aligner-eval>.

Table 3

Automatic alignment quality with both the default and custom configurations of Bertalign (Bertalign_d and Bertalign_c, respectively) and the Galechurch length-based algorithm.

	Bertalign_c	Bertalign_d	Galechurch
P	0.888	0.862	0.427
R	0.905	0.857	0.368
F1	0.896	0.859	0.396

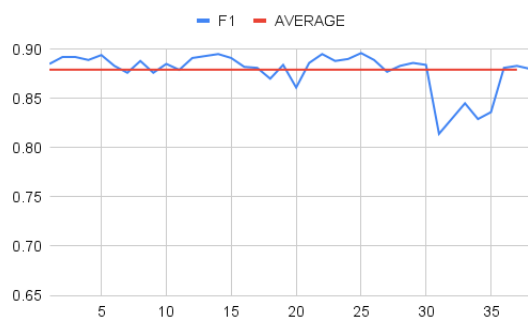


Figure 3: F1 score (on the y axis) by chapter (on the x axis): to facilitate the reading of the chart, the vertical axis has been set to 0.65. The average line is displayed in red.

is not great, with an average F1 of 0.879. However a drop can be noted in the range between chapters 31 and 35 which describe the plague in Milan with numerous historical digressions, often not translated. In particular, chapters 31 and 32 of the original text are merged into a single chapter in the translation in which there is a high number of omissions covering 33% of all the alignments. In addition, that group of chapters includes cross-alignments that are not correctly handled by Bertalign. On the contrary, the best F1 (0.896) is found for chapter 25 in which 1:1 alignments, the simplest type, are 73% of the total.

5. Conclusion and Future Work

This paper described the creation of a parallel corpus aligned at sentence level made of the whole text of Ventisettana, that is the version of the novel published by Manzoni in 1827, and the 1834 anonymous English translation. This resource is made available on Github in XLM format¹² and will be also uploaded in the ILC4CLARIN repository. The whole aligned corpus has been used as gold standard for evaluating Bertalign, an embedding-based automatic sentence aligner. Results obtained with

¹²https://github.com/RacheleSprugnoli/Sentence_Alignment_Manzoni.

a custom setting of the parameters are compared to the ones achieved with the default options and with a length-based algorithm (Galechurch) showing very good performances, with an F1 slightly below 0.9.

The activity presented here served as a laboratory for future experiments which will concern the other editions of the novel and the main translations into neo-Romance languages. In particular, a sentence level alignment activity of chapter VIII is underway taking into account the largest possible number of available English translations also considering, thanks to an agreement with the translator, the very recent American translation of the novel [27]. The choice of maintaining the sentence unity in the Italian text will facilitate the comparison between different translations and, consequently, investigations on the choices made by the translator in a diachronic perspective.

The alignment at the word level of some chapters of the Ventisettana with the English edition of 1834, already adopted for the sentence level alignment, is also in progress. In this case, the alignment is done using Ugarit [28].¹³ Unlike what has been done in other projects [29], in our project the aim of the alignment does not concern the creation of a translation memory for machine translation purposes, but the analysis of the choices made by the translator: for this reason, the alignment is performed considering punctuation and also between linguistic elements whose literal correspondence is rather fuzzy. This choice makes it possible to highlight oversights, errors and singular innovations of the translator. The output of our manual alignment will be used to evaluate automatic approaches, such as fast_align¹⁴ and AWESOME¹⁵.

Acknowledgments

Questa pubblicazione è stata realizzata da ricercatrice con contratto di ricerca cofinanziato dall'Unione europea - PON Ricerca e Innovazione 2014-2020 ai sensi dell'art. 24, comma 3, lett. a, della Legge 30 dicembre 2010, n. 240 e s.m.i. e del D.M. 10 agosto 2021 n. 1062. Questa ricerca è stata anche finanziata dall'Università degli Studi di Parma attraverso l'azione Bando di Ateneo 2022 per la ricerca co-finanziata dal MUR-Ministero dell'Università e della Ricerca - D.M. 737/2021 - PNR - PNRR - NextGenerationEU.

References

- [1] P. Bellezza, *Il Manzoni all'estero*, in: *Curiosità manzoniane*, Vallardi, Milano, 1923, pp. 57-73.

¹³<https://ugarit.ialigner.com>.

¹⁴https://github.com/clab/fast_align.

¹⁵<https://github.com/neulab/awesome-align>.

- [2] P. Bellezza, *Attraverso le traduzioni dei "Promessi sposi"*, in: *Curiosita' manzoniane*, Vallardi, Milano, 1923, pp. 75–95.
- [3] G. Getto, *Manzoni europeo*, Ugo Mursia, Milano, 1971.
- [4] P. Frare, *Manzoni europeo?*, *Nuovi quaderni del Crier*. I "Promessi sposi" nell'Europa romantica 9 (2012) 199–220.
- [5] V. Intonti, R. Mallardi, *Cultures in contact: Translation and reception of i promessi sposi in 19th century england*, 2011.
- [6] T. R. Steiner, *English translation theory 1650-1800*, 2, Rodopi, 1975.
- [7] M.-A. Lefer, *Parallel corpora*, in: *A practical handbook of corpus linguistics*, Springer, 2021, pp. 257–282.
- [8] J. Tiedemann, *Parallel data, tools and interfaces in OPUS*, in: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 2214–2218. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf.
- [9] R. Steinberger, M. Ebrahim, A. Poulis, M. Carrasco-Benitez, P. Schlüter, M. Przybyszewski, S. Gilbro, *An overview of the european union's highly multilingual parallel corpora*, *Language resources and evaluation* 48 (2014) 679–707.
- [10] Y. Xu, A. Max, F. Yvon, *Sentence alignment for literary texts: The state-of-the-art and beyond*, in: *Linguistic Issues in Language Technology*, Volume 12, 2015-Literature Lifts up Computational Linguistics, 2015.
- [11] P. F. Brown, J. C. Lai, R. L. Mercer, *Aligning sentences in parallel corpora*, in: *29th Annual Meeting of the Association for Computational Linguistics*, 1991, pp. 169–176.
- [12] W. A. Gale, K. W. Church, et al., *A program for aligning sentences in bilingual corpora*, *Computational linguistics* 19 (1994) 75–102.
- [13] D. Varga, P. Halácsy, A. Kornai, V. Nagy, L. Németh, V. Trón, *Parallel corpora for medium density languages*, volume 292, Amsterdam; Philadelphia; J. Benjamins Pub. Co, 2007, p. 247.
- [14] M. Kay, M. Roschisen, *Text-translation alignment*, *Computational linguistics* 19 (1993) 121–142.
- [15] R. Sennrich, M. Volk, *Mt-based sentence alignment for ocr-generated parallel texts*, in: *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas: Research Papers*, 2010.
- [16] B. Thompson, P. Koehn, *Vecalign: Improved sentence alignment in linear time and space*, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 1342–1348. URL: <https://www.aclweb.org/anthology/D19-1136>. doi:10.18653/v1/D19-1136.
- [17] L. Liu, M. Zhu, *Bertalign: Improved word embedding-based sentence alignment for chinese-english parallel corpora of literary texts*, *Digital Scholarship in the Humanities* 38 (2023) 621–634.
- [18] E. Signoroni, *Evaluating the state-of-the-art sentence alignment system on literary texts.*, in: *Recent Advances in Slavonic Natural Language Processing (RASLAN 2021)*, 2021, pp. 115–124.
- [19] G. Raboni, *«manzonionline»: considerazioni in corso d'opera*, *Griseldaonline* 20 (2021) 149–155.
- [20] Z. S. Villar, O. A. Pinedo, *Taligner 3.0: A tool to create parallel and multilingual corpora*, in: *Corpora in Translation and Contrastive Research in the Digital Age: Recent advances and explorations*, John Benjamins, 2021, pp. 125–146.
- [21] R. Sprugnoli, A. Redaelli, M. Sartor, *Risorse linguistiche per lo studio dei "Promessi sposi"*, in: *La memoria digitale. Forme del testo e organizzazione della conoscenza. Atti del XII convegno annuale AIUCD (Siena, 5-7 giugno 2023)*, Università degli Studi di Siena, Siena, 2023, pp. 301–303.
- [22] P. Vondřička, *Aligning parallel texts with InterText*, in: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association (ELRA), Reykjavik, Iceland, 2014, pp. 1875–1879. URL: http://www.lrec-conf.org/proceedings/lrec2014/pdf/285_Paper.pdf.
- [23] M. Straka, *Udpipe 2.0 prototype at conll 2018 ud shared task*, in: *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 2018, pp. 197–207.
- [24] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, *Stanza: A Python natural language processing toolkit for many human languages*, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2020. URL: <https://nlp.stanford.edu/pubs/qi2020stanza.pdf>.
- [25] J. Munday, S. R. Pinto, J. Blakesley, *Introducing translation studies: Theories and applications*, Routledge, 2022.
- [26] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, W. Wang, *Language-agnostic BERT sentence embedding*, in: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 878–891. URL: <https://aclanthology.org/2022.acl-long.62>. doi:10.18653/v1/2022.acl-long.62.
- [27] A. Manzoni, *The Betrothed: A Novel*, *Modern Li-*

- brary, 2022. Translated by Michael F. Moore.
- [28] T. Yousef, C. Palladino, F. Shamsian, M. Foradi, Translation alignment with ugarit, *Information* 13 (2022) 65.
- [29] T. Yousef, C. Palladino, F. Shamsian, A. d. Ferreira, M. F. dos Reis, An automatic model and gold standard for translation alignment of ancient greek, in: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 5894–5905.