

# Digital Explorations of Historical Multilingual Practices. The Challenges of the HyperAzpilcueta Project

Manuela Bragagnolo<sup>1</sup>, Marcus Pöckelmann<sup>2</sup>, Polina Solonets<sup>3</sup>, Andreas Wagner<sup>4</sup>

<sup>1</sup> University of Trento, Italy – manuela.bragagnolo@unitn.it

<sup>2</sup> Martin Luther University Halle-Wittenberg, Germany – marcus.poeckelmann@informatik.uni-halle.de

<sup>3</sup> Max Planck Institute for Legal History and Legal Theory, Germany – solonets@lhlt.mpg.de

<sup>4</sup> Max Planck Institute for Legal History and Legal Theory, Germany – wagner@lhlt.mpg.de

## ABSTRACT (ENGLISH)

Based on the example of an early modern case of self-translation and collaborative translation, the contribution seeks to open discussion about the relation between multilingual practices embodied in textual research objects on the one side and the “user experience” of digital platforms and their affordances on the other.

**Keywords:** text collation; digital multilingual practices; HyperAzpilcueta

## ABSTRACT (ITALIANO)

*Esplorazioni digitali di pratiche multi linguistiche storiche. Le sfide del progetto HyperAzpilcueta*

Basandosi sull'esempio di un caso di autotraduzione e traduzione collaborativa della prima età moderna, il contributo cerca di aprire una discussione sulla relazione tra le pratiche multilinguistiche incarnate negli oggetti di ricerca testuale, da un lato, e la “user experience” delle piattaforme digitali e delle loro possibilità, dall'altro.

**Parole chiave:** collazione di testi; pratiche digitali multilinguistiche; HyperAzpilcueta

## 1. INTRODUCTION

Historical multilingual practices involve much more than mere linguistic translation. For this reason, conceptualizing the main aspects of those practices and their application in different textual witnesses is a challenge in itself. But the question then is: how can digital environments and methods facilitate investigation of the various aspects of these practices and their differences?

As an illustration of the general problem, we present “HyperAzpilcueta”, a digital synoptic edition of several printed editions of an early modern bestseller, *Manual de Confessores* (Manual for Confessors). Its author, the Spanish canon law professor Martín de Azpilcueta (1492-1586), translated and simultaneously substantially revised his work several times, incorporating different types of authorities, examples and arguments in the different editions. Together with his printers, he even organized typographical devices for his respective target audience to better orient themselves in the text. With the help of translations and revisions, he thus devised new ways of producing different types of knowledge of (religious, spiritual, but also legal) normativity.

## 2. MARTÍN DE AZPILCUETA’S MANUAL DE CONFESORES AS A CASE OF MULTILINGUAL PRACTICE

First printed in Portugal in 1549, the *Manual* was then transformed, updated, and translated into different languages by Azpilcueta himself, in a revision process that lasted for almost 40 years. Four of the editions showcasing the most important transformations are used in the HyperAzpilcueta project: the first edition, printed in Coimbra in Portuguese in 1549; the second Portuguese edition, with major transformations (Coimbra, Portuguese, 1552); the 1556 Spanish translation by Azpilcueta, largely enriched (Salamanca, 1556); and finally, the Latin translation, completely revised in structure (Rome, 1573).

As Azpilcueta was working on translation alone or in cooperation with other scholars and practitioners of the time, his work on the *Manual* is an example of both, “collaborative translation” (Cordingley & Frigau Manning, 2017) and “self-translation” practices (Miglietti, 2024). Such a multilingual practice was quite common in the early modern world, with every new edition used to update, restructure, and adapt the text (and the book as a material object) to new target audiences, new contexts of production and circulation, and to changes in societal practices that form the subject matter of many parts of the book.

In what dimensions can we study the translation practice of the *Manual*?

Firstly, the differences between the texts reflect the varying contexts of writing, book production and dissemination. They actually “accompanied” Azpilcueta’s moving from Portugal (Coimbra, 1549; Coimbra,

1552) to Spain (Salamanca, 1556), and from Spain to Rome (Rome, 1573). Every new edition and “self-” or “collaborative” translation was produced in close contact with a different set of local actors (Bragagnolo, 2022; Bragagnolo, 2024): throughout his career, missionaries and confessors have contacted Azpilcueta with questions and doubts about the latest (or earlier) edition(s) of the *Manual*, as did kings and penitents, theologians and jurists, not to forget book printers and traders. Economic and political constellations varied locally and changed substantially during the 40 years of revising the *Manual*, and the Council of Trent (1545–1563) revised in important ways the official Catholic doctrine on many questions treated in it. Therefore, each new version incorporated the feedback of a particular community of practice in its revised and translated language.

Secondly, the different versions of the *Manual* aimed to reach different reader groups: The first edition (in 8° pragmatic vernacular handbook) was meant for simple and unlearned confessors, active in remote parishes of Portuguese provinces. By contrast, the very erudite, rich, in 4° Latin *Enchiridion* was intended to reach learned confessors and missionaries of the universal Church, widespread all over the world, via academic channels of communication and book distribution. The editions in between reflected this path in their language, from the vernacular to Latin, with injections of doctrine, and increasingly technical, complex and erudite use of language.

Those differences between contexts of production and target audiences found expression in concrete linguistic phenomena: different audiences demand different sets of references to authorities, different contexts provide different sets of examples (mundane or contrived, colonial or courtly), different languages have different proverbs and personae. And in terms of knowledge organization, Azpilcueta developed different ways of mapping and navigating the subject matter: even though the basic structure of 27 chapters was kept intact in all editions, each edition shows another ordering of arguments, examples and references, mostly with new sections added. Different structural and typographical finding devices such as sub-chapter headings and summaria, the wrapping of newly added passages in asterisks or their introduction by daggers etc. suggest very different uses of the book by the readers. Some of them, such as the marginal numbers, even seem to indicate an (again, multilingual) practice of navigating across and comparing the different editions on the part of the readers.

### 3. CONCEPTUALIZING THE COMPREHENSION OF MULTILINGUAL PRACTICE IN A DIGITAL PLATFORM

The question arises of how a digital platform can facilitate in the study of the various described aspects of the *Manual* and texts with a similar history of tradition. The following section discusses the main requirements and the possibilities already offered by existing approaches.

First, it is essential to allow researchers to closely examine the text itself by presenting it in full detail whenever necessary (close-reading). This should also take into account – may even by optional highlighting – their various special elements such as different fonts, marginal numbers, literature references or special characters (e.g. asterisks and daggers in our case). If the texts are encoded in TEI XML, this is supported by many tools for generating digital editions such as TEI-Publisher,<sup>1</sup> EVT<sup>2</sup> (Rosselli Del Turco et al., 2014) or the Versioning Machine<sup>3</sup> (Schreibman et al., 2003).

A second essential feature is the possibility to compare certain passages of the text witnesses with each other in order to allow researchers to identify similarities and textual variants. Typical close-reading representations for this purpose in collation tools highlight or arrange the words in specific ways. This includes classic synoptic views, which can be found in the already mentioned tools (TEI-Publisher, EVT and Versioning Machine) as well as in LERA (Pöckelmann et al., 2022), or variant graphs as in CollateX (Dekker et al., 2015) or TRAViz (Jänicke et al., 2015). However, the underlying collation algorithms in current tools are typically designed for witnesses of the same language and rather simple textual variants like the replacement of individual words or phrases with alternatives. While research for handling multiple languages in the domain of scholarly editions is presently going on (Balck et al., 2025; Reboul, 2024; Hohenegger et al., 2023), available collation tools still lack the ability to automatically point the researcher to semantic similarities and differences on a word or phrase level when it comes to multilingual texts. When dealing with larger amounts of text like the *Manual* has, the collation in hierarchical steps is reasonable (Pöckelmann et al., 2022), i.e., in a first step larger segments of the text such as sentences or

---

<sup>1</sup> <https://teipublisher.com/exist/apps/tei-publisher-home/index.html> (cons. 05.01.2025).

<sup>2</sup> <http://evt.labcd.unipi.it> (cons. 05.01.2025).

<sup>3</sup> <http://v-machine.org> (cons. 05.01.2025).

paragraphs are aligned before in a second step the collation on the word level is performed. This presupposes the constitution and alignment of those segments in the first place, which becomes harder the more substantial the revisions of the text witnesses are. Besides the presentation and user interface, features of automatic segmentation and detection of semantic similar phrases or sentences that differ only on the surface, i.e., mostly in a changed wording but not meaning, are relevant for the platforms or toolchains we are interested in. Promising recent solutions outside the domain of scholarly editing are based on word or sentence embeddings (e.g. Kusner et al., 2015). While these approaches have been intended for monolingual applications at first, they were soon be extended to the multilingual realm (Reimers & Gurevych, 2020; Ham & Kim, 2021), resulting in state-of-art approaches for effective sentence-alignment even across languages, e.g. SentAlign (Steingrímsson et al., 2023). However, these have difficulties when dealing with major structural differences among the text witnesses, as it is the case with the *Manual*. Another problem lies in the training of the underlying embeddings, which requires a large amount of digitized text. This is simply not available as we face so-called *low-resource languages* in this project due to the historical state and the specific genre. In our alternative approach to embeddings, we apply a rule-based system for segmentation with an alignment algorithm that profits from the similarity of the languages in the *Manual* (Wagner & Bragagnolo, 2019) in combination with an evaluation of some of the particular features mentioned above and are currently extending the collation platform LERA<sup>4</sup> (Pöckelmann et al., 2022) accordingly.

Even if the used algorithms deliver good results, it is unlikely that any automatic segmentation or alignment will be perfect at extracting and evaluating all the relevant features nor that the results correspond exactly to the researcher's expectations. This weakness is amplified by the fact that often changes in one aspect (like segmentation) entail substantial further changes in another (alignment). Thus, a digital platform must provide means for researchers to correct both segmentation and alignment in the later close reading process. It should be possible to join or split segments, move them around (in the sense of adjusting their position in the synoptic view), and even place copies of existing segments – marked as such – at other positions to reflect transpositions.

Finally, the question arises as to how the entirety of the work and the identified revisions can be presented. With the large volume of text on the one hand and the structural changes in terms of both sequencing and volume on the other, there is a need for overview visualizations (distant-reading) in order to navigate through these volumes of text as well as to compare structural changes. The latter may include transpositions that span over long distances within the different text witnesses. A recent survey (Yousef & Jänicke, 2021) lists a huge variety of visualization methods that has been developed over the years to represent alignments, some of which are applicable to huge text witnesses for distant-reading. Furthermore, some older approaches have focused on conveying the dynamics of revision processes or structural aspects (e.g. Fry, 2009; Cheesman & Roos, 2017). However, due to the combination of various aspects of the *Manual* – multilinguality, restructuring and the mere existence of more than two witnesses – no existing visualization is sufficient. A flexible, zoomable, or elastic way of providing overviews of different levels of distance seems desirable (as in Bludau et al., 2020; Pöckelmann et al., 2024). This could include the text itself at the lowest zoom level, making it a scalable-reading approach.

#### 4. CONCLUSION

As we have shown, for many of the mentioned requirements, some functionalities are, in principle, available in existing tools. Yet the combination of tasks (represent structural as well as semantic changes, establish alignment, facilitate close reading, process etc.), the volume of text to be processed and presented and the multitude of features to take into account – and the multilinguality affecting all the tasks individually and in combination – mean that there is, as of now, no platform that really does facilitate study of historical multilingual practices like we have presented with Apilcueta's *Manual*. Moreover, to the best of our knowledge, for the issue of distinguishing "salient" from "trivial" differences between texts in different languages (where practically everything is different in terms of linguistic surface forms), no solution at all is available. The state of affairs rather seems to demand a discussion about how to conceptualize "salient semantic similarity or difference" in a multilingual context, how to anchor any mechanism in this regard to concrete words or phrases in the source documents, and (only then) how to convey eventually acquired measurements in a user interface.

---

<sup>4</sup> <https://lera.uzi.uni-halle.de> (cons. 05.01.2025).

## REFERENCES

- Balck, S., Dähne, J., Etling, F., Fischer, F., Frenzel, S., Grote, B., Heße, S., Molitor, P., Pöckelmann, M., Ritter, J., Singh, Y. & Stede, M. (2025). *Collation of Multilingual Versions of a Text: Necessity, Approach, Challenges* [Conference presentation abstract]. ADHO DH2025. Lisbon; Portugal.
- Bludau, M.-J., Brüggemann, V., Busch, A. & Dörk, M. (2020). Reading Traces: Scalable Exploration in Elastic Visualizations of Cultural Heritage Data. *Computer Graphics Forum*, 39(3), 77-87. <https://doi.org/10.1111/cgf.13964>.
- Bragagnolo, M. (2022). Les voyages du droit du Portugal à Rome. Le Manual de confesseurs de Martín de Azpilcueta (1492-1586) et ses traductions. In J.-L. Fournel, I. Paccagnella (Eds.), *Traduire. Tradurre. Translating. Vie des mots et voies des œuvres dans l'Europe de la Renaissance* (pp. 209-228), Droz.
- Bragagnolo, M. (Ed.). (2024). *The Production of Knowledge of Normativity in the Age of the Printing Press. Martín de Azpilcueta's Manual de Confessores in global perspective*, Brill. <https://doi.org/10.1163/9789004687042>
- Cheesman, T. & Roos, A., (2017). Version Variation Visualization (VVV): Case Studies on the Hebrew Haggadah in English. In M. Büchler & L. Mellerin (Eds.), *Computer-Aided Processing of Intertextuality in Ancient Languages* [Special Issue]. *Journal of Data Mining & Digital Humanities*. <https://doi.org/10.46298/jdmdh.3704>
- Cordingley, A. & Frigau Manning, C. (Eds.) (2017). *Collaborative Translation: from the Renaissance to the Digital Age*, Bloomsbury Academic. <https://doi.org/10.5040/9781350006034>
- Dekker, R., Van Hulle, D., Middell, G., Neyt, V. & Van Zundert, J. (2015). Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project. *Digital Scholarship in the Humanities*, 30(3), 452-470. <https://doi.org/10.1093/llc/fqu007>
- Fry, B. (2009) *On the Origin of Species: The Preservation of Favoured Traces*. <https://benfry.com/traces/>
- Ham, J., Kim, E.-S. (2021). Semantic alignment with calibrated similarity for multilingual sentence embedding. In M.-F. Moens, X. Huang, L. Specia & S. W. Yih (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2021* (pp. 1781-1791). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.findings-emnlp.153>
- Fry, B. (2009) *On the Origin of Species: The Preservation of Favoured Traces*. <https://benfry.com/traces/>
- Hohenegger, H., Montemagni, S., Pasini, E., Ciotti, F., Mancinelli, T., De Longis, E., Chiaravalli, I., Boschetti, F., Del Grosso, A. M. & Meschini, F. (2023) *DiScEPT – Digital Scholarly Editions Platform and aligned Translations*. <https://www.studigermanici.it/progetto-discept>
- Jänicke, S., Geßner, A., Franzini, G., Terras, M., Mahony, S. & Scheuermann, G. (2015). TRAViz: A Visualization for Variant Graphs. *Digital Scholarship in the Humanities*, 30(suppl\_1), i83–i99. <https://doi.org/10.1093/llc/fqv049>
- Kusner, M. J., Sun, Y., Kolkin, N. I. & Weinberger, K. Q. (2015). From Word Embeddings to Document Distances. In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. (pp. 957-966). <https://dl.acm.org/doi/10.5555/3045118.3045221>
- Miglietti, S. (2024). Translations of self-translations in early modern Europe: preliminary remarks and exploratory case studies. *Philological Quarterly*, 102(2-3), 199-223.
- Pöckelmann, M., Damm, A. & Rebigier, B. (2024). *An Interactive Visualization of Recurring Microforms Across Various Hebrew Works and their Witnesses* [Conference presentation abstract]. DH2024, 34. international annual conference of Digital Humanities, Washington D.C., United States.
- Pöckelmann, M., Medek, A., Ritter, J. & Molitor, P. (2022). LERA—an interactive platform for synoptical representations of multiple text witnesses. *Digital Scholarship in the Humanities*, 38(1), 330–346. <https://doi.org/10.1093/llc/fqac021>
- Reboul, M. (2024). Alignement sémantique et manque de données : l'apport des modèles de langue. Le cas du latin et du grec. In B. Bordalejo, R. Risam & E. Château-Dutier (Eds.), *DH Unbound 2022, Selected Papers* [Special Issue]. *Digital Studies/Le champ numérique* 13(3): 1–13. <https://doi.org/10.16995/dscn.9669>.
- Reimers, N. & Gurevych, I. (2020). Making monolingual sentence embeddings multilingual using knowledge distillation. In B. Webber, T. Cohn, Y. He & Y. Liu. (Eds.), *Proceedings of the 2020*

- Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 4512-4525). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.365>
- Rosselli Del Turco, R., Buomprisco, G., Di Pietro, C., Kenny, J., Masotti, R. & Pugliese, J. (2014). Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions. *Journal of the Text Encoding Initiative*, 8. <https://doi.org/10.4000/jtei.1077>
- Schreibman, S., Kumar, A. & McDonald, J. (2003). The versioning machine. *Literary and Linguistic Computing*, 18(1), 101–107. <https://doi.org/10.1093/lc/18.1.101>
- Steingrímsson, S., Loftsson, H. & Way, A. (2023). SentAlign: Accurate and scalable sentence alignment. In Y. Feng & E. Lefever (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 256-263) <https://doi.org/10.18653/v1/2023.emnlp-demo.22>
- Wagner, A. & Bragagnolo, M. (2019). *Multimodale Versuche der Alignierung historischer Texte*. [Conference presentation abstract] DHd2019, 6. Jahrestagung der Digital Humanities im deutschsprachigen Raum. Frankfurt, Germany, 181–184. <https://doi.org/10.5281/zenodo.4622224>.
- Yousef, Y. & Jänicke, S. (2021). A Survey of Text Alignment Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1149–1159. <https://doi.org/10.1109/TVCG.2020.3028975>