

A Multimodal Mapping Methodology for Translators. How to quickly familiarize yourself with a subject area terminology

Sílvia Araújo, Micaela Aguiar ^{1,2}

^{1,2} *University of Minho, Rua da Universidade, Braga, 4710-057, Portugal*

Abstract

In this paper, we propose a methodology for multimodal mapping of a subject area terminology, aimed at the development of multiskills in translator training. The methodology follows the three phases of the translation process: pre-translation (compilation of a parallel specialized corpus), translation (assisted and automatic with CAT tools) and post-translation (creation of different visualizations of single and multi-words extracted from the corpus). It is intended to help translators quickly familiarize themselves with a subject area terminology by going beyond translation, i.e., creating multimodal content with the potential to simplify and communicate science.

Keywords

Multimodal Mapping, Terminology, Translation, Information visualization, Digital literacy

1. Introduction

The modern society's demand for more and more specialization has had a profound impact on the Humanities, as they seek to be competitive in today's markets [1]. In this context, we propose a methodology focused on translator autonomy, which consists in appropriating a subject area terminology — the first level of contact with a specialized language — by carrying out DIY corpora activities, building translation memories and extracting terminology. This methodology fosters multimodal and creative training, by suggesting different ways of visualizing information, such as word clouds and digital glossaries, thus meeting the markets' demand that translators take a more active role in multilingual content creation. Moreover, the creation of visualizations helps to improve terminological understanding of a subject area and further contributes to making scientific knowledge more accessible, i.e., achieving a democracy of data [2].

2. Multimodal Mapping of a subject area terminology


This methodology is designed in three phases that accompany the translation process: pre-translation (compilation of DIY corpora), translation (translation activity with CAT tools) and we propose a third step that goes beyond translation — post-translation (creation of multimodal byproducts). This methodology was implemented as part of the Corpus Linguistics course of the Master's in Translation and Multilingual Communication (2020-2021). The tasks were carried out collaboratively, in groups of three students. The source of the corpora were the Portuguese and English versions of the MSD manuals for professionals, which specialize in medical information. MSD Manuals are a comprehen-

Words/Machines #4: Text Simplification and Adaptation, March 25, 2022, Brest Cedex, France

EMAIL: saraujo@elach.uminho.pt (A. 1); maguiar60@gmail.com (A. 2)

ORCID: 0000-0003-4321-4511 (A. 1); 0000-0002-5923-9257 (A. 2)



!  2022 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)

sive source of medical information that includes thousands of subjects in all areas of medicine. They are offered as a free public service to health professionals and the general public [3].

2.1. Task 1 - Pre-Translation: Compiling a Parallel Corpus

As Wilkinson [4] states, freely available corpora of Language for Specific Purposes do not abound, therefore it is essential that translators are able to create their own specialized corpora specifically designed for their needs. Thus, creating DIY corpora should be part of an active pedagogical practice for translators-in-training, as it fosters proactivity of future translators, equipping them with the tools to create useful resources in preparation for the actual translation phase. The first task consisted in compiling a parallel Portuguese-English corpus of two medical subjects. The students collected the entries of each medical subject of the Portuguese and English versions in a Notepad++ .txt document, then proceeded to automatically align the entries at the sentence level using the LF Aligner tool [<https://sourceforge.net/projects/aligner/>], and finally reviewed the alignment process manually. The result was a bitext, i.e., an .xls document with texts in both languages, later used for terminology extraction.

2.1.1.Sub-task 1: Translation Memory

The students created a translation memory using LF Aligner to convert the .xls document into .tmx format. Translation memories are an essential tool for translators, especially as the need increases to translate large volumes of specialized texts in ever tighter deadlines. According to Reinke [5], translation memories allow the translator “to recycle these translated segments by automatically proposing a relevant translation from the memory as a complete (“exact match”) or partial solution (“fuzzy match”) whenever the same or a similar sentence occurs again” (p. 55-56). Reinke concludes by remarking that translation memories increase translators' productivity, consistency, and speed. The translation memory created by the students was later used in the translation activity.

2.1.2.Sub-Task 2: Extraction of Terminology

In this task, the students extracted the 100 most frequent single words and the 100 most frequent multi-words in Portuguese and in English from each of the two chosen medical subjects. The terminology extraction was carried out automatically using Sketch Engine's One Click Terms extractor [<https://terms.sketchengine.eu/>], and after that students manually reviewed the results of the automatic extraction.

2.2.Task 2 - Translation Activity with CAT tools

For this task, the students imported the translation memory into MemoQ, and proceeded with the assisted and automatic translation of one of the entries coming from the MSD Manuals. This mixed approach, which combines computer-aided translation and machine translation, is relevant because the use of machine translation is nowadays very common in translation companies, namely in multinational companies with a very large workflow. Considering the increasing ubiquity of machine translation systems, to a point where their benefits will make them indispensable, it is important to develop post-editing skills in translators [6].

2.3.Task 3 - Post-Translation: Visualization for Science Communication

In this stage, students were challenged to go beyond the translation phase and consider ways in which the results from the pre-translation phase could be used to improve their understanding of the subject area whilst creating visual and appealing resources for science communication and populariza-

tion², such as word clouds and glossaries³. The purpose of this task was to provide students with a transversal, multimodal education that promotes creativity and digital literacy [7, 8].

2.3.1.Sub-task 3: Word Clouds

Word clouds are a visualization format of the most frequent terms in a document, which can serve as a means of summarizing texts [9] and/or as a starting point for deeper analyses [10]. For this reason, we believe word clouds offer a dynamic and interesting way of getting to know the terminology of a given subject area. The students used the lists of the 100 most frequent single word and multi-words extracted in the pre-translation task and selected only the specialized terms, excluding more generic terms, such as *patient* or *symptoms*. This sorting process contributes to greater appropriation of the subject area terminology. From the resulting list, they created single-word and multi-word clouds, in Portuguese (Figure 1) and in English (Figure 2).



Figure 1: Example of Specialized Multi-Word Cloud in Portuguese

² As part of a related initiative, we have been collaborating with the Medical School of University of Minho in training young doctors to use medical terminology to create informative brochures (click [here](#) to see an example in Portuguese). These brochures, in addition to having an accessible definition of the disease, present more frequent and accessible terms to describe the symptoms and risk factors, i.e., the terminology appears in context and adapted to the target audience.

³ For example, the science communication project *Frontiers for Young Minds*, which features young people as reviewers in the peer review process, has adopted precisely this idea of placing under each article a glossary with the terms considered most difficult (example: <https://kids.frontiersin.org/articles/10.3389/frym.2021.751592>).

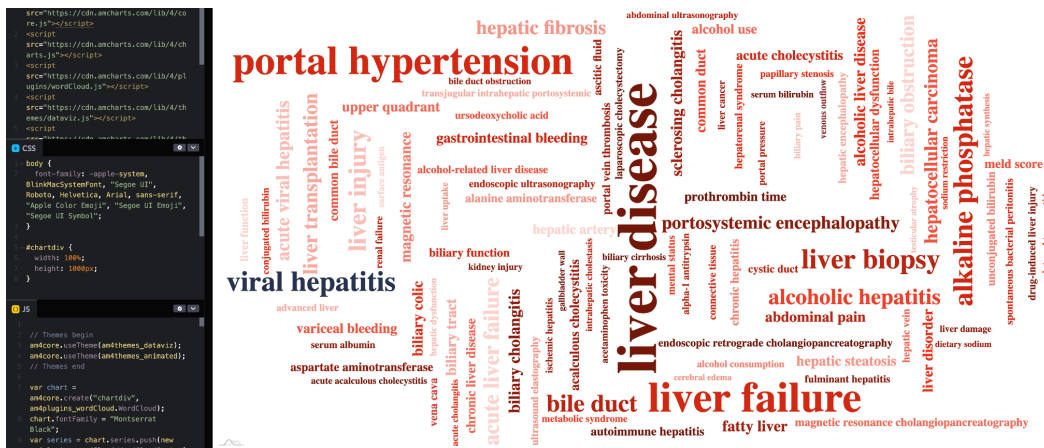


Figure 2: Example of Specialized Multi-Word Cloud in English

The word clouds were created using a script that was provided to the students (see figures 1 and 2 above), since, in general, word cloud generators available online (such as tagcrowd [https://tagcrowd.com/], tagxedo [https://tagxedo.com/], wordart [https://wordart.com/], wordcloudmaker [http://wordcloudmaker.com/], among others) work at the single-word level, and therefore lexical units formed by two or more graphic words are lost in the clouds. The clouds created with the script are interactive.

2.3.2.Sub-task 3: Digital Bilingual Glossaries

In this task, each group member selected 10 Portuguese terms from the previously extracted list of the 100 most frequent terms to create a glossary, with a total of 30 terms. For each term, students retrieved the definition, the genre, and the English equivalent. After compiling the information, the students were provided with a script which they used to create a digital XML-annotated glossary that each group could customize using CSS (Figure 3). These glossaries can contain images, videos and sound related to the concepts. Through this task students are introduced to annotation and programming languages, essential for the development of localization skills and multilingual digital production.



Figure 3: Example of Digital Bilingual Glossary

3. Conclusion

In this paper, we have proposed a methodology for multimodal mapping of a subject area terminology (Figure 4), which fits into a holistic education for translators, focused on the development of multiskills. The multimodal mapping methodology was designed to encompass the stages of the translation process and suggests a third step that goes beyond the translation stage: the creation of multimodal content.

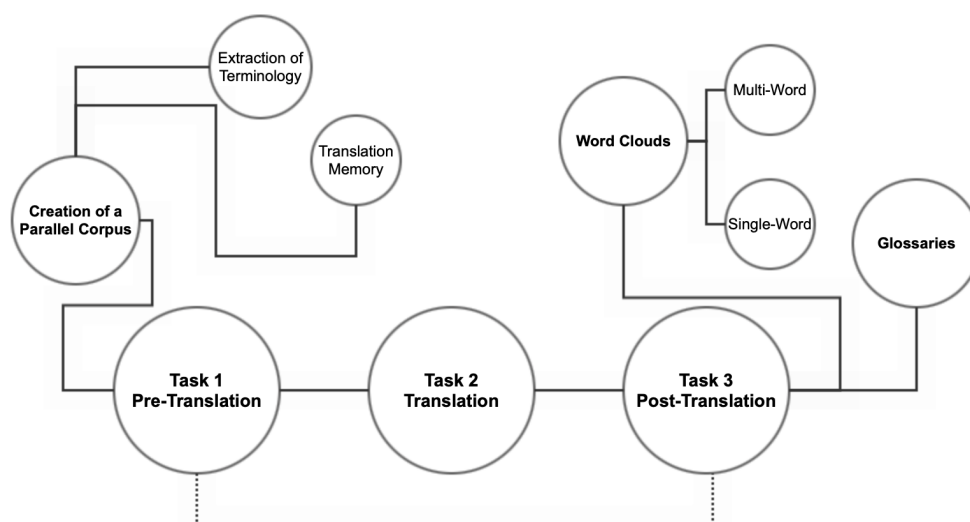


Figure 4: Multimodal Mapping Methodology for Translators

This methodology focuses on promoting translator autonomy by providing them with the skills to create DIY corpora, automatically extract terminology and create translation memories, before the actual translation stage. As the digital literacy [11, 12] component of the multimodal mapping methodology, students are asked to mobilize different digital tools (such as LF Aligner, MemoQ, Sketch Engine, Notepad++, among others) and to get to know programming and annotation languages (such as XML and HTML/CSS). In addition, students use the substrate of the pre-translation stage to create multimodal contents adapted to different target audiences (see the example of medical brochures in footnote 2) and/or to specific communicative goals, e.g. science simplification and communication, brand internalization, cultural dissemination, etc. This approach highlights the relevance of training multi-faceted translators who are able to create multimodal, multilingual and inclusive content aimed at contributing to the construction of an informed society that is prepared to overcome the communicative barriers imposed by specialised languages. The methodological proposal described in this paper stems from work originally developed in the PortLinguE project, a funded R&D project for the creation of an online portal for specialized languages [13].

3. Acknowledgements

The PortLinguE project (ref. PTDC/LLT-LIG/31113/2017) is funded by ERDF under Portugal 2020 and by national funds through the Portuguese Foundation for Science and Technology.

4. References

1. S. Maruenda-Bataller, & B. Clavel-Arroitia, (Eds.) Multiple Voices in Academic and Professional Discourse: Current Issues in Specialised Language Research, Teaching and New Technologies, Cambridge Scholars Publishing, 2011.

2. P. Mirowski, The future(s) of open science, *Social Studies of Science*, 48(2), pp.171–203, 2018. doi:10.1177/0306312718772086
3. MSD Manuals. URL: <https://www.msmanuals.com/en-pt>
4. M. Wilkinson, Compiling Corpora for Use as Translation Resources, *Translation Journal*, 10(1), 2006. URL: <http://translationjournal.net/journal/35corpus.htm>
5. U. Reinke, State of the art in Translation Memory Technology, in: G. Rehm, F. Sasaki, D. Stein & A. Witt (Eds.), *Language technologies for a multilingual Europe: TC3 III (Translation and Multilingual Natural Language Processing 5)*, Language Science Press, pp. 55–84, 2018. URL: <https://www.blogs.uni-mainz.de/fb06-tc3/files/2015/11/25-145-3-PB.pdf>
6. J. Nitzke, S. Hansen-Schirra, *A short guide to post-editing. (Translation and Multilingual Natural Language Processing 16)*. Berlin: Language Science Press, 2021. doi: 10.5281/zenodo.5646896
7. G. Damaskinidis, The Visual Aspect of Translation Training in Multimodal Texts. *Meta* 61, n° 2, pp. 299–319, 2016. doi:10.7202/1037761ar
8. M. Boria, A. Carreres, M. Noriega-Sánchez, M. Tomalin (Eds.), *Translation and multimodality. Beyond words*, London: Routledge, 2020.
9. K. Padmanandam, S. P. V. D. S. Bheri, L. Vegesna, & K. Sruthi, A Speech Recognized Dynamic Word Cloud Visualization for Text Summarization, in: *6th International Conference on Inventive Computation Technologies (ICICT)*, pp. 609–613, 2021. doi:10.1109/ICICT50816.2021.9358693
10. F. Heimerl, S. Lohmann, S. Lange, & T. Ertl, Word Cloud Explorer: Text Analytics Based on Word Clouds, in *47th Hawaii International Conference on System Sciences*, pp. 1833–1842, 2014. doi:10.1109/HICSS.2014.231.
11. D. Kergel, M. Paulsen, J. Garsdal, & B. Heidkamp-Kergel, *Bildung in the Digital Age: Exploring Bildung through Digital Media in Education (1st ed.)*, 2022. Routledge. doi:10.4324/9781003158851
12. M. Anstey, G. Bull, *Foundations of Multiliteracies. Reading, Writing and Talking in the 21st Century*, Routledge, 2018.
13. S. Araújo, C. Analide, A. Silva; A. Correia, P. Ferreira (forthcoming), *PortLinguE: breve apresentação de um projeto para Línguas de Especialidade e comunicação de ciência [PortLinguE: short presentation of a project for Specialised Languages and science communication]*. *Revista Cadernos BAD*.