

Software localisation to help terminological accuracy

Vitor Gonçalves

Centro de Investigação em Educação Básica (CIEB)
Instituto Politécnico de Bragança
Bragança, Portugal
vg@ipb.pt

Bruno F. Gonçalves

Centro de Investigação em Educação Básica (CIEB)
Instituto Politécnico de Bragança
Bragança, Portugal
bruno.goncalves@ipb.pt

Abstract — Software localization is the process of adapting a software, web or mobile app to the culture and language of users in a specific market. Is a growing phenomenon, resulting from the pressing need that various public and private entities must adapt their products and services to one or more languages, locations and/or cultures. In this context, the study presented herein had as its main objective to analyse and compare two tools (OpenOffice Writer and LibreOffice Writer) in their English version and their respective translation into Portuguese, with the aim of understanding their level of terminological quality based on the Portuguese language. What is the degree of consistency between the terminology used in the interface of each of the two tools, corresponded to the main research question. The comparative study seemed to us to be the research methodology that would best suit an analysis of concrete data and respective deduction of similarities and divergences of the elements. Inconsistency can lead to incomprehension, less speed in handling and, potentially, constant errors on the part of those using the software were some of the main results. It is also pointed out points that are considered important for the improvement of localisation support tools, as well as possible suggestions to facilitate the improvement of terminological accuracy in the final products. It is hoped that this study may instigate more attentive and detailed studies that contribute to the improvement of the localisation of such tools.

Keywords - CAT tools; comparison study; localisation; terminological accuracy; translation.

I. INTRODUCTION

This paper corresponds to a first comparison study within the scope of Software and Web Page Localisation, in the context of translation processes. It is important to stress that, since this is a comparative and quantitative study, it does not propose solutions or even a qualitative analysis, i.e., it does not intend to attribute a degree of quality or present scenarios inherent to quality control processes.

Therefore, it was intended to compare the localisation of two applications to assess the level of consistency in terms of terminology, in this case the terminology used in the toolbars of two Word processors. Terminology is the study of, and the field of activity concerned with the collection, description, processing, and presentation of terms, i.e., lexical items belonging to specialised areas of usage of one or more languages [1]. Localisation is, according Schäler [2], “the linguistic and cultural adaptation of digital content to the requirements and the locale of a foreign market; it includes the provision of services and technologies for the management of multilingualism across the digital global information flow”. Thus, localisation activities include translation (of digital material as diverse as user

assistance, websites, and videogames) and a wide range of additional activities. Contrary to definitions provided by the Localisation Industry Standards Association (LISA), this definition explicitly focuses on digital content and includes the management of multilingualism as one of the important localisation activities.

In other words, we have on the one hand a growing phenomenon, that of localisation, which is inserted in the context of an increasingly globalized world, intertwined with a clear need for companies wishing to expand into new markets. Essentially, localisation refers to the process of adapting a product or service to meet the needs and expectations of a specific target audience in a specific place. This includes translating text, adapting images and videos, adapting user interface elements and tailoring promotion and marketing messages to be relevant and understandable to local consumers.

Localisation is important because it helps make products and services accessible and attractive to consumers in different parts of the world, "a modest price to pay to enter international markets" (p. 147) [3]. This is especially important for companies that want to expand their business in other countries or regions. Localisation can also be important for companies that wish to serve consumers in different languages or cultures, as it can help avoid miscommunication and increase customer satisfaction. In addition, localisation can be an important factor in a consumer's purchasing decision and can help foster customer loyalty. Terminology is the set of specific terms used in a particular field or subject matter. This includes technical, scientific, legal, business terms, etc. This requires standardization of the process of compiling, describing, processing, and presenting the terms with the aim of optimizing communication among specialists and professionals concerned with the standardization of language. Terminology is important because it helps to ensure that messages are conveyed accurately and consistently across different languages and contexts.

In translation, terminology is especially important because it helps to ensure that terms are translated accurately and consistently. This is important because many terms have specific and precise meanings in different languages, and an error in translation can lead to inefficient and sometimes linguistically incorrect communication or even serious problems, such as medical or legal errors. Furthermore, terminology is important for maintaining consistency and accuracy of messages throughout a translation project, ensuring that terms are translated consistently throughout the document or project.

This paper therefore intends to present the results of the research undertaken. Begin by presenting the work methodology and carry out the theoretical framework cementing the importance of the central issue of consistency at the terminology level. Next, it presents the data analysis and proceed to their discussion. Finally, it include the main conclusions and final considerations, as well as possible future work.

II. METHODOLOGY APPROACH

According to [4], "the comparative method consists of investigating things or facts and explaining them according to their similarities and differences. Generally, the comparative method approaches two series or facts of analogous nature, made of social means or of another area of knowledge, to detect what is common to both" (p. 40).

This study corresponds to the comparison of two or more corpora to discover a relationship of similarity between the two. It carried out a comparative analysis between two Word processors, specifically OpenOffice Writer and LibreOffice Writer. It was decided to employ the comparative model according to Williams and Chesterman [5].

Any research makes use of a theoretical model of the object being studied, either explicitly or implicitly. So, if we are studying translation, or the translating process, we need some preliminary model of this kind to orient ourselves, to give ourselves an initial framework within which we can begin to think. What exactly is a model? A model is a construction that represents some aspect of reality. Think of a miniature scale model of a car, representing a full-size car: it looks like a copy of it, but it is many times smaller [5].

In this model the work consisted in comparing the source text (ST) and the target text (TT), and in our case, in comparing the source texts of two different tools and the target text for two different tools. In this paper, the focus was mainly on the analysis of the target texts, i.e., on the terms already translated into Portuguese checking if there was similarity with the terms used in the target language, English.

Our research is essentially empirical (and not conceptual) and quantitative. Empirical research seeks new data, new information derived from the observation of data and from experimental work; it seeks evidence which supports or disconfirms hypotheses or generates new ones [5]. In quantitative research, the aim is to be able to say something about the generality of a given phenomenon or feature, about how typical or widespread it is, how much of it there is; about regularities, tendencies, frequencies, distributions. Ultimately, quantitative research may aim at making claims about universality. Quantitative research seeks to measure things, to count, and to compare statistically. (...) It is often said that qualitative research is more subjective, and quantitative research more objective [5].

We questioned whether there was consistency in terms of the terminology used.

The beginning went through the analysis of the functionalities of both processors and recording them on a spreadsheet. Then it was needed to enter these tables into a Computer Assisted Translation (CAT) tool to create bilingual

tables and then enter the translations. Once this was done, it was necessary to analyze the two documents to see where there were differences in the terms used and, then, calculate their degree of consistency as a percentage.

III. TERMINOLOGICAL CONSISTENCY

Terminological consistency is important because it helps ensure that messages are conveyed accurately and understandably to the intended audience. When terms are used consistently in a text or document, it helps avoid confusion or ambiguity and ensures that the reader understands what is being said.

In addition, terminological consistency is important for maintaining the accuracy and quality of the text or document. When terms are used consistently, it helps to ensure that the meanings of terms are clear and precise throughout the document, which is especially important in technical or scientific texts. Terminological consistency is also important for maintaining the cohesion and fluency of the text. When terms are used consistently, it helps create a sense of unity and cohesion in the text, which can help improve readability and reader comprehension.

Terminological consistency is important for the image of the company or organisation. When terms are used consistently across all company communications, this can help create an image of professionalism and credibility, which can help attract and retain customers.

Terminology errors in translation can have various consequences, depending on the context in which they occur and the severity of the errors. Some possible consequences of terminology errors in translation include:

- Confusion or misinterpretation: If terms are translated incorrectly or inconsistently, this can lead to confusion or misinterpretation on the part of readers, listeners or users. This can affect the effectiveness of communication and can lead to misunderstandings or problems.

- Reputational damage: Terminology errors can negatively affect the reputation of a company or organisation. If terms are translated incorrectly or inconsistently, this can create an impression of a lack of professionalism or a lack of care for the quality of communication.

- Legal problems: In some cases, terminology errors in translation can lead to legal problems. For example, if a term is translated incorrectly in a contract or legal document, this can affect the validity or interpretation of the document.

- Loss of customers: Errors in terminology can lead to customer dissatisfaction or even the loss of the business relationship with those same customers. If terms are translated incorrectly or inconsistently, this can affect the customer's trust in the company or organisation and can lead to cancellation of orders or termination of the business relationship.

IV. THE IMPORTANCE OF TRANSLATION SUPPORT TOOLS

To ensure consistency, it is standard for translators or localisers to use support programmes and CAT tools. CAT tools for translation and software localization nowadays offer, almost

without exception, a bundle of functions that tend to include alignment, concordancing and term extraction already described, and quality assurance and even workflow management described later in this chapter. The notion of an integrated suite of tools is captured by the common label 'workstation'. This section focuses on the distinctive features of translation and localization tools, as well as tools for subtitling, that support the translator's core task [6].

These tools make use of two key resources: translation memories (TM) and terminology databases (TD). TMs are tools used to help ensure terminological consistency and quality in a translation project. They store the terms and phrases used in a translation project so that translators can access and use them consistently throughout the project. MTs can also store examples of usage of the terms and phrases, to help translators understand how they should be used in different contexts. They are usually created and managed by a translation project manager or terminology specialist (in the case of larger projects, or by individual translators when they are not part of a larger team) and can be shared with translators to help them understand how terms should be used in the project. In addition, MTs can be updated and expanded during the project to include new terms or phrases that may be needed. TD are collections of terms and definitions used in a specific field or subject matter. TD can also be used to help ensure terminological consistency and quality in a translation project by providing translators and localizers with access to accurate and consistent information about the terms used in the project. Terminology databases can be created manually or using specialized terminology management tools. They can be shared with translators or other users via online or software platforms. In addition, terminology databases can be integrated with machine translation (MT) tools, CAT tools, or translation project management tools to help ensure translation consistency and accuracy.

V. DATA ANALYSIS AND DISCUSSION

With the analysis of the terms acquired from the two Word processors used to conduct this research, a total of 907 terms (79%) were surveyed from OpenOffice Writer and LibreOffice Writer, with 234 (21%) of them repeated in the graphical user interface (GUI) in the toolbars of the two programs. These 907 terms represent only a small portion of all the terminology of these programs, since the objective of this study was a general analysis of the interfaces, i.e., only of the toolbars that constitute a core element of the GUI, and not an in-depth analysis of all the constituent elements of that interface. We thus arrived at a sample of 907 terms taken from the two programs.

Considering the 234 repeated terms, 202 (86%) show terminological consistency, but 32 (14%) of them do not show terminological consistency. The remaining 673 terms out of the 907 total terms, are different terms between the two Word processors. It should also be noted that, since the aim was to compare all instances of consistency, the focus of this comparative analysis was not on those 673 terms that are not repeated in our sample. This is since these 673 terms, not being repeated between the two tools, are not comparable, since for this study the important thing is comparing equal terms in the two TPs.

The analysis concluded that there is a high degree of terminological consistency between the two Word processors used to carry out this research, since it is easy to see that there is a considerable majority (86%) of terms that present consistency or a high degree of consistency. High degree can be in cases, such as:

- Automatic summary of presentation - Automatic summary for presentation.

- Capital letter at the beginning of the sentence - Capital letter at the beginning of the sentence.

Even in cases where there is no consistency, often this inconsistency does not represent major problems for the user. Since there is a possibility that many localizations have been done by localization teams with different variants of Portuguese, namely European Portuguese and Brazilian Portuguese. Let's see for example:

- Global document template - Master document.

- Total pages - Count pages.

In short, we have a total of 907 terms raised between the two programmes with a total of 234 terminological repetitions between their visual interfaces, of these repetitions we found 32 different and 202 equal with an inconsistency level of 13.6%.

After data analysis, it is possible to say there is 86.4% consistency of terms between the locations of the two tools. This study is empirical and quantitative, as the aim was not to draw illations about the quality of the translation itself, but to make the quantification of the terminological metrics of two similar types of word processors.

VI. CONCLUSION

In the last decade, the strategy of globalisation has become crucial in the digital society in which we are now living [7]. Today, more than ever, the software localisation industry assumes an importance that institutions and companies should not neglect [8].

Assuming the software localization is the process of adapting a software, web or mobile app to the culture and language of users in a specific market [9], the focus of this comparative study was on checking terminological consistency between OpenOffice Writer and LibreOffice Writer and not as a way of providing solutions in obtaining it or even for quality assurance of these tools. With the rigorous analysis carried out, we have concluded that there is 86% terminological consistency. This aspect shows us that consistency is something that is present in the minds of the companies that created the word processors, and simultaneously, in those of the localisers.

Any type or degree of inconsistency found was ultimately inconsequential and we hypothesise it to have been fundamentally the result of one of two things: (i) there may not have been any contact between the teams or individuals who made the different localisations of the two word processors and as such there will not be, or is not expected to be, TD sharing; (ii) there will not even be a TD in use as a result of no CAT tool being employed to provide TD or MT creation.

We hope that the results of this study may instigate other researchers to conduct further study to provide reliable solutions to obtain terminological consistency as well as quality control, which may benefit the localisation and software and website creation industries.

The literature made available in the scope of Terminology and Terminography and Software or Web Pages Localisation has served to awaken interest in two areas that may seem distinct at first, but naturally have strong links and appear to be one of the areas with the greatest potential for future growth.

ACKNOWLEDGMENT

This work has been supported by FCT – Fundação para a Ciência e a Tecnologia within the Project Scope: UIDB/05777/2020.

This paper was based in study works of the curricular unit of Software Localization of the Translation Master (IPB).

REFERENCES

- [1] J. C. Sager, *A practical course in terminology processing*, Amsterdam, Netherlands: John Benjamins Publishing Company, 1990.
- [2] R. Schäler, "Localization and translation" in *Handbook of translation studies*, vol. 1, Y. Gambier and L. Van Doorslaer, Eds. John Benjamins Publishing, 2010.
- [3] K. Dunne, "Localization and the (R)evolution of Translation" in *A companion to translation studies*, S. Bermann and C. Porter, Eds. Wiley-Blackwell, 2014.
- [4] O. Fachin, "Fundamentos de metodologia", 4.ed., São Paulo: Saraiva, 2005.
- [5] J. Williams and A. Chesterman, *The map: A beginner's guide to doing research in translation studies*. Manchester, UK: St Jerome Publishing, 2002.
- [6] J. Munday, "Introducing Translation Studies, Theories and Applications", *Babel*, vol. 55:3, pp. 288-301, 2009.
- [7] Yunker, *Beyond Borders: Web Globalization Strategies*, Indianapolis: New Riders, 2014.
- [8] A. Brashi, "An Adaptive Methodology to Overcome Localization Translation Challenges", *International Journal of English Linguistics*, 11(4), 105, 2021.
- [9] Phrase., *The Complete Guide to Software Localization*, 2023, Available in <https://phrase.com/blog/posts/software-localization/>
- [10] F. Jahn, E. Ammenwerth, V. Dornauer, K. Höffner, M. Bindel, T. Karopka, & A. Winter, "A Linked Open Data-Based Terminology to Describe Libre/Free and Open-source Software: Incremental Development Study". *JMIR medical informatics*, 11, e38861, 2023. <https://doi.org/10.2196/38861>