The Impact of Translation Memory Segmentation on the Cohesion and Coherence of Technical Texts Translated from English into Arabic

Amira Mahmoud Muhammad Abdel-maksoud

An instructor at the AUC, Cairo University (CHP) and (CLT)
A PhD Candidate,
English Department, Faculty of Arts
Cairo University

Abstract:

Even though Translation Memories (TMs) are known to increase translation productivity, term consistency and quality, many scholars pinpoint that TM segmentation disrupts the Source Text (ST) linearity. Being an indispensable tool for translators and a novel tool in the Arab world, there is a growing need for an in-depth investigation of this feature. This research investigates the impact of TM segmentation on the Target Text (TT) in terms of cohesion and coherence. It is a mixed research design where participants translate two English user manuals into Arabic. The participants translated the first English smart phone user manuals into Arabic without the software program WordFast Anywhere (WFA) and translated the second through WFA. Three professional translation instructors at the AUC Language Department (AUC LD) assessed the language of the two texts produced in terms of cohesion and coherence based on the analytical rubric Farahzad (1992) and analyzed the protocols produced. The evaluation of pre-test, post-test and the Think-Aloud Protocols(TAP) revealed that TM segmentation negatively affects the cohesion and coherence of the TT.

Keywords: Translation memory, segmentation, cohesion, coherence, cognitive activity, CAT tools, WordFast Anywhere.

(ISSN : 2536 - 9555)

تأثير تقسيم ذاكرة الترجمة للنص المترجم إلى وحدات على التماسك النصي للنصوص التقنية المترجمة من الإنجليزية إلى العربية

الباحثة / أميرة محمود محمد عبد المقصود

مدرس بالجامعة الأمريكية ومحاضر ببرنامج الساعات المعتمدة (قسم لغة إنجليزية – جامعة القاهرة) ومركز اللغات الأجنبية والترجمة التخصصية بجامعة القاهرة

طالبة دكتوراة - قسم اللغة الإنجليزية – كلية الآداب – جامعة القاهرة

ملخص البحث باللغة العربية:

رغم أن برامج ذاكرة الترجمة تشتهر بقدرتها على زيادة الإنتاجية والحفاظ على ثبات المصطلحات والإرتقاء بالجودة أثناء الترجمة، يؤكد كثير من الباحثين أن خاصية تقسيم النص المترجم منه إلى "وحدات ترجمة" يتسبب في كسر التسلسل الخطي للنص المترجم منه. ولإن ذاكرة الترجمة أحد الأدوات التي لا يمكن للمترجم في العصر الحالي الاستعاضة عنها في جميع أنحاء العالم وكونها أحدى الأدوات الحديثة التي بزغ إستخدامها في العالم العربي، هإن هناك حاجة متزايدة لتناولها بمزيد من العمق. لذا، فإن هذا البحث يهدف إلى سبر أغوار خاصية تقسيم النص المترجم منه إلى وحدات وأثرها على الحبك والسبك في النص المترجم إليه. تستخدم الدراسة الحالية برنامج " Wordfast Anywhere" كنموذج لبرامج ذاكرة الترجمة. يقوم المشاركون في الدراسة بترجمة دليلي استخدام هاتف نقال؛ احدهما باستخدام الحاسب الألي بدون استخدام أي برامج ذاكرة ترجمة والأخر باستخدام برنامج ذاكرة الترجمة. يتولى عدد من مدرسي الترجمة المتخصصين بقسم اللغات بالجامعة الأمريكية بتقييم التراجم من حيث السبك والحبك بالاستناد المعيار القياسي "Farahzad Rubric" المستخدم في تقييم النصوص المترجمة والذي يدرج الحبك والسبك ضمن أحد اهم عناصر تقييم النص المترجم.

الكلمات الدالة : تكنولوجيا الترجمة، ذاكرة الترجمة ،التماسك النصي ، دراسات الترجمة، برنامج وورد فاست اني وير " WFA"

1. Introduction

Regardless of the pivotal role TMs have researchers such as Krüger (2016) believes that the way literature tackles the use of TM systems is neither explicit nor structured. Furthermore, Mor (2018) confirms the need for more empirical research studying translators' interaction with TMs. Bundgaard, Christensen, and Schjoldager (2010) advocate approaching CAT tools from a novel dimension rather than the heavily investigated dimensions, namely, speed, accuracy, and consistency. Therefore, the present study seeks to unravel the impact of using the software program WordFast Anywhere (WFA) to translate an English user manual into Arabic. It specifically studies the effect of the segmentation feature on the TT in terms of cohesion and coherence. It adopts a product-oriented approach as it assesses the language of the TTs produced in terms of cohesion and coherence based on Farahzad (1992). It is an analytical rubric that is suitable for evaluating long texts as the case in the present study. The present study uses two smart phone user English manuals as the ST and considers the Arabic translation the TT. One manual will be translated as a whole unit without being segmented and the other will be uploaded and translated on WFA in the form of segments.

2. Statement and Rationale of the Problem

Due to the integral role technology plays in translation industry and the rapid changes this field witnesses, translation memories have become an indispensable element in the translators' workstation worldwide. They are extensively investigated by many scholars including Candel-Mora (2015). However, this investigation is limited in number and scope. It targets language pairs rather than English and Arabic and is limited to exploring the detrimental impact TMs have on creativity, remuneration, and translators' status. Furthermore, the little literature investigating TM segmentation covers the translators' perception only. It does not investigate TM segmentation as a default feature that disrupts

(ISSN : 2536 - 9555)

the whole-unit text unity of the ST (Teixeira & O'Brien, 2017). It does not take into account quality of the final product.

In light of the above-mentioned, Christensen and Schjoldar (2010) and Mor (2018), state that the impact segmentation has on the final text in terms of cohesion and coherence is a gap in literature that requires further investigation. Mahfouz (2018) states that TMs are given scant attention in empirical studies specifically in the Arab world where the English/Arabic language pair is prevalent in the translation market. Consequently, the mainspring for conducting the current empirical research is to elaborate on the segmentation feature imposed by TMs and its impact on the (TT) produced in terms of cohesion and coherence.

3. Research Question

The present study seeks to answer the following question:

a) What is the difference between smart phone user manuals translated from English into Arabic when the ST is unsegmented and when the ST is segmented in terms of textual cohesion and coherence of the TT?

4. Literature Review

4.1Definition of Translation Memories and WFA.

Bowker (2002) defined TM as a linguistic database that can be used to store and retrieve STs and their corresponding TTs so that translators can reuse them when a new similar ST is encountered. Reinke (2013) defined TM as a special word processing package that translators seek support from during the translation process. As for Wordfast Anywhere (WFA), the prime focus of the current thesis, it is the online version of the popular WordFast CAT program. It is a distinguished standalone cloud-based TM which enables translators to process data on remote servers and this obviates the need to purchase expensive computers.

4.2History of CAT Tools and Translation Memories.

Although TMs came into being in the 1970s, they were not extensively used till the nineties (Garcia, 2014). By mid 1990s,

more advanced systems started to depend on translation memory, terminology management, alignment tools, and more advanced features. Since that time, TMs' benefits captured the attention of a considerable number of scholars and a considerable amount of literature on the significant role TMs play in the translation process was published. This included Alotaibi (2014), Bowker (2002), Killman (2018), Le Blanc (2013), Pym (2011) and Vela et al., (2019). All these studies deemed TMs the most widely used translation tool in the market because they save time and accelerate the pace of the translation process. There was a consensus among them all that CAT tools are the core and essence of all modern translations carried out. Regardless of the added value these researches represented to literature, their contribution was limited to unveiling the benefits TMs encompass while the drawbacks that these programs may entail were hardly discussed.

4.3 Typologies, Creation and Utilization of Translation Memories

Gambín (2014) divided TMs into two main categories, desktop-based TMs and cloud-based TMs, namely. As for desktop-based TM, they are software programs which are purchased from software companies and installed on PCs. They are used off-line through a license the translation memory company provides the users with. Although desktop-based TMs do not require connectivity to a server, they necessitate providing enough space on the user's PC.

Reviewing literature comparing Desktop-based TMs to cloud-based TMs, Šanca (2018) preferred the new generation of TM systems where all the operations are done online to desktop-based TMs. Users of cloud-based TMs just create an account on the software program through the email and set a personal password to be able to use the program and the options it offers. Chan (2017) believed that these tools are ascribed this status because they improve the translational turnaround time compared to desktop-based TM; they allow multiple translators to work on the same translation project simultaneously and allow reviewers to work on

(ISSN : 2536 - 9555)

a document that is still being translated. Finally, Pym (2013) preferred cloud-based TMs than desktop-based TMs. He believed that the more TMs are in the cloud, the more they become accessible to users and the more they will be used.

Even though cloud-based TM software programs were privileged by all the above scholars and researchers than desktop-based TMs, Florez and Alcina (2015) and Moreno (2018) stated that they were understudied. The available discussions and evaluations of cloud-based TMs were limited to the point of view of the researchers and developers while the user's point of view was neglected.

As for TM creation, Kumar (2018) confirmed that users can have a TM through three different methods: creating an empty TM at the beginning of the translation process; importing a populated one from another TM database or compiling one through the alignment process. As regards utilization of TMs, Garcia (2014) stated that TMs are mainly used in translation of non-literary texts due to the repetitive nature of these texts. TM segmentation.

According to Killman (2018) and Mor (2018) segmentation is one of the five main features any TM system has. Users of desk-top based TMs and cloud-based TMs receive the ST divided into segments by default. According to Garcia (2014) a segment is a single sentence or phrase demarcated with segment terminators or typographic marks of the text such as a full stop or a question mark. A segment can be a smaller unit such as a heading, a list, a bullet point, a caption or content of a cell. Kumar (2018) distinguished the term segment form other similar terms such as translation unit (TU). To illustrate, a TU is the translated segment stored by the TM system in the ST and the TT to be reused in case of repetition.

Literature on TM features showed contentious opinions about TM segmentation. Esselink (2000), for example, presented a neutral opinion about segmentation describing it as a main process and a key concept in all TM systems. Austermühl (2006), however, described it as an obstacle that hinders the way to creativity and

innovation. De La Cova (2016) explained that segmentation compels TM users to work with isolated texts in sequential order.

Apropos the literature tackling the segmentation feature, much criticism is levelled to it. For instance, Bowker (2015), and Christensen, Schjoldager, and Flanagan (2017) affirmed that literature on TM segmentation was neither explicit nor structured. They are limited to mere description and opinions without any empirical evidence.

4.4 TMs in Relation to Cohesion and Coherence.

According to Halliday and Hasan (1976) cohesion and coherence are the essence of textuality. They described coherence as the embodiment of the unity exhibited on the grammatical and semantic levels. They believed that the relationship between cohesion and coherence is a causative one where cohesion leads to coherence. Diao (2016) and Malamkjaer (2017), however, acknowledged the close interrelationship between cohesion and considered them an inseparable factor when assessing a text. They ascribed this stance to the spread of the elements of cohesion including appropriate use of pronouns, linkages, transitional words, and coherence all over the text.

4.5 Technical Texts

A technical text is one of the five text types translators encounter (Hatim and Mason, 1996). A technical text is any text that deals with a specific knowledge such as instruction guides, installation guides and software user manuals (Byrne, 2006). Alaoui (2015) classified manual user guides under technical non-literary texts which is void of any literary tendencies. It is not read for the purpose of learning more about a certain subject nor intended to entertain or impress specific readers.

According to Byrne (2006) software user guides are drafted by one of the technical writers working for the software company within the framework of developing and selling the product as part of the product development process. Ehrensberger-Dow (2019)

(ISSN : 2536 - 9555)

believed that translating manuals necessitates discourse awareness. To produce high-quality TT that maintains the specialized subject matter, the terminology and linguistic complexity these manuals include, the translator should be fully aware of the ST discourse. All these precautions add to the constraints imposed on the translator and make user manuals a rich text for researchers.

4.6 Previous Studies

Vela, et al., (2019) conducted a study on the server-based TM software program CATaLog. The study included 16 undergraduate students who were enrolled on a translation technology class within the framework of a TS program. The study targeted evaluating the usability of this cloud-based TM software program from the human post-editing performance and the efficiency of the solutions proposed by TM and MT. The participants were asked to post-edit 30 news items translated from English into German. The results revealed that the participants preferred CATaLog online; they liked the output of the MT system and the color scheme of the suggestions proposed by CATaLog rather than the one proposed by MateCAT software program.

Furthermore, Bundgaard and Christensen (2019) conducted a study on TMs at workplace. They wanted to ensure that the data gathered on the usability of TMs and their effect on the translator's cognitive activity is ecologically valid. The results showed that the concordance search was the translators' preferred resource. At the end of the study, Bundgaard and Christensen (2019) stressed the need for further research on how professional translators interact with CAT tools, and highlighted the need to prepare translation students for the future. They also concluded that the other TM features merit further attention and considered these features another possible area of future research.

Mor (2018) studied effect of TMs on the phenomenon of linguistic interference. The study included 90 participants divided into three groups who worked on an unpopulated TM. The results reported that TMs may negatively affect cohesion of the TT rendered. Mor (2018) attributed this potential negative impact to

the avoidance of using anaphoric or cataphoric references on the part of the translators to be able to find matches easier. Although Mor addressed a gap literature, included a relatively large number of participants and used various data-gathering methods, it reveals controversial results as the tools used were varied in number and order and the participants had various profiles.

To further study the challenges which translators face when TMs are utilized, De La Cova (2016) conducted a preliminary qualitative research on a localization project of a web application. The research scope encompassed all the constraints imposed by the localization process in general including the segmentation feature as well. The corpus examined was limited to one single ad hoc monolingual English corpus that included the user interface strings of three web applications. This study was a real contribution to literature on TM segmentation, notwithstanding, it was limited to analyzing the English ST.

Candel-Mora (2015) explored the potential textual constraints TM segmentation is likely to pose. He attempted to verify the accusation leveled at TM tools claiming that TMs have negative effect on TT. He used a small ad hoc comparable corpus of many press releases in English that were rendered into Spanish. The findings revealed that TM tools have some constraints on the quality of the TT in terms of "coherence, cohesion, ortho-typography, anaphoric and cataphoric references, linearity, readability" (p. 73). However, the corpus analysis was described by Candel-Mora at the end of the study as a limitation. He further added that corpus analysis was an inappropriate tool that hindered the generalization of the results found about cohesive ties or textual cohesion.

Turning to the Arab world, the literature revealed a very limited number of researches which is ascribed to the novelty of TMs (Mahfouz, 2018; Alotaibi, 2014).

Mahfouz (2018) tried to explore users' attitudes to CAT tools in the Arab world through a survey and semi-structured interviews.

(ISSN : 2536 - 9555)

She investigated the attitudes of 114 translation students and professional translators in Egypt. The research related the users' attitudes towards CAT tools to various factors including years of experience, computer skills and type of texts translated. She also used semi-structured interviews for collecting data and this increased the research's strength through the use of the mixed-method that ensures reliability of results. The study revealed a favorable attitude among participants towards using CAT tools regardless of the mixed and contradicting opinions it revealed on some aspects. Mahfouz 2018 asserted that the better computer skills the users have, the more favorable attitudes towards CAT tool users adopt. Based on the literature on TMs in the Arab world, this study is considered a contribution. Conversely, the methodology adopted was limited to surveys and semi-structured interviews. This study would have been more effective in case TMs had been studied from a different perspective. Another limitation was the sampling as including both translation students and professional translators made the sample loses it harmonious nature. Such limitation could have been the reason behind reaching conflicting and controversial opinions. Mahfouz (2018) concluded her study by describing TMs as promising avenue for Arab researchers to explore and asserted that there is a desperate need in the Arab world for empirical research on TMs and their usability.

Accordingly, the current study aims to fill a gap in the literature. It targets investigating the effect TM segmentation may have on the translator's cognitive activities when translating technical texts from English into Arabic. It seeks providing more empirical research on the actual interaction between translators and TM systems in practical contexts along with exploring TM's usability from the perspective of practicing translators which is still required as confirmed by Krüger (2016). Embracing a multi-dimension perspective, the current study integrates a process-oriented approach and a product-oriented approach. The process-oriented approach is represented by the Think Aloud Protocol TAP

and the product-oriented approach is exemplified in assessing the TTs produced in terms of cohesion and coherence.

5. Methodology

The curent study adopted a one group pre-test post-test design. The participants received and translated two texts from English into Arabic, one as a pre-test and the second as a post-test. The TT rendered when the pre-test is one-unit and the TT rendered when the post-test was segmented by the TM are evaluated by three different experienced translation instructors based on Farahzad rubric to overcome data subjectivity. The current research adopted convenience sampling which by recruiting all the Egyptian male and female students enrolled in a 6-week translation course at the AUC where the researcher works. The pre-test is extracted from a 167-page user manual for smart phone "Galaxy Note 10". The participants rendered four pages (420 word) into Arabic without the TM software. As for the post-test (430 word), it is an extract from a 188-page user manual for smart phone “Samsung Note 20” of which three pages, 50 segments, were rendered under the effect of TM segmentation. The first procedure for data collection took place in session (6) when the participants rendered the pre-test and was followed by rendering the post-test in session (7). Before the pre-test was administered, all the participants received a clear explanation about the purpose of the research. The translations provided through the pre-test and the post-test were collected and evaluated by three different translation instructors. Then, the cohesion and coherence score each participant had in the pre-test and the post-test based on Farahzad rubric were compared.

(ISSN : 2536 - 9555)

6. Results

To answer the research question, the pre-test and the post-test are graded through Farahzad rubric (1992) based on a 20-point scale by three different translation instructors. The 20 points which are assigned to transitional words, appropriate use of pronouns, linkages, pronoun references and natural flow of ideas.

In conclusion, a paired-samples t.test conducted to compare the grades given on cohesion and coherence when students translate a one-unit ST compared to the grades the students received when translating under the effect of TM segmentation. There is a significant difference in the scores achieved when the ST is not segmented (mean=16.40, standard deviation-= 0.81) and the score achieved when the ST is segmented (mean=11.94, standard deviation-= 1.54) conditions; t (59) = -21.4188, p=.000. The difference in means is statistically significantly (difference = -4.46). The results suggest that students' grades were significantly lower when TM is utilized than when the TM segmentation is not used.

7. Limitations and Recommendations for Future Research

The current study has some limitations. Two of the participants show higher performance due to previous experience in the translation field. Tackling novice translators entailed some proficiency problems. A deeper analysis using a larger sample with a variety of dimension could confirm the results or produce more insight. Another study can explore the same topic using professional translators. Also, the current study focuses on the performance of novice Egyptian translators using English as the SL and Arabic as the TL.

References

Alaoui, A. (2015) Knowledge transfer and the translation of technical texts. *International Journal of Humanities and Social Sciences,9(10).3380-3386*

Alotaibi, H. (2014). Teaching CAT Tools to Translation Students: an Examination of Their Expectations and Attitudes, *Arab world English journal Special*, (3) 65-74

AUSTERMÜHL, F. (2006). Training translators to localize. In A. Pym, A. Perekrestenko & B. Starink (Eds.) *Translation Technology and its Teaching* (pp. 69-82). Tarragona.

Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. University of Ottawa Press.

Bowker, L. (2015). *Terminology and translation*. *Handbook of terminology*. John Benjamins.

Bundgaard, K., & Christensen, T. (2019) (n.d.). Is the concordance feature the new black? A workplace study of translators' interaction with translation resources while post-editing TM and MT matches. *Journal of Specialized Translation,31*(31),14-37

Bundgaard, K. & Christensen, T. & Schjoldager, A. (2016). Translator-computer interaction in action — an observational process study of computer-aided translation. *The Journal of Specialized Translation, 25*, 106-130.

Byrne, J. (2006). *Technical translation: usability strategies for translating technical*

documentation. Dordrecht: Springer.

Chan, S. (2017). *The future of translation technology: Towards a world without Babel*. Routledge.

Candel-Mora, M. (2015) Comparable corpus approach to explore the influence of computer-assisted translation systems on

(ISSN : 2536 - 9555)

textuality. *Prodceia-Social Behavioral Sciences*, *198(2015),* 67-73

Christensen, T. & Schjoldager, A. (2010). Translation-memory (TM) research: What do we know and how do we know it? *HERMES - Journal of Language and Communication in Business*, (44), 89–101.

Christensen, T., Flanagan, M., & Schjoldager, A. (2017). Mapping translation technology research in translation studies. An introduction to the thematic section. *HERMES - Journal of Language and Communication in Business*, (56), 7- 20. https://doi.org/10.7146/hjlcb.v0i56.97199

De la Cova, E. (2016). Translation Challenges in the Localization of Web Applications. *Sendebar*, *27*, 235-266

Diao, Y. (2016). Coherence in translation: from reader's cognition-with ulysses as case study. *Theory and Practice in Language Studies*, *6*(2), 33- 336. http://dx.doi.org/10.17507/tpls.0602.14

Ehrensberger-Dow, M. (2019). Ergonomics and the translation process. *Slovo.ru: Baltic Accent,10*(1), 37-51

Esselink, B. (2000). *A practical guide to localization*. John Benjamins

Florez, S. & Alcina, A. (2015) Quality model for the evaluation of open translation technologies. In P. Sandrini & M.García,(Eds.), *Translation and openness* (pp.81-114).Innsbruck University Press.

Garcia, I. (2014). Computer-aided translation. In S. Chan. (Ed.), *The Routledge encyclopedia of translation technology* (pp. 68-87). Routledge

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Hatim, B., & Mason, I. (1996). *The translator as communicator*. Routledge.

Killman, J. (2018). A Context-based approach to introducing translation memory in translator training. In B. Godev (Ed.) *Translation, globalization and translocation* (pp137– 159). Palgrave/Macmillan.

Krüger, R. (2016) Contextualising computer-assisted translation tools and modelling their usability. T*rans-kom 9* (1), 114-148

Kumar, A. (2018). Structuring of Translation Memory Structuring of Translation Memory Work-Integrated Learning Programmes Division.

LeBlanc, M. (2013). Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies. *The International Journal of Translation and Interpreting Research, 5* (2), 1–13.

Mahfouz, I. (2018). Attitudes to CAT Tools: Application on Egyptian Translation Students and Professionals. *Arab World English Journal*, *4*(4), 69-83.

Malmkjaer, K. (2017). *The Routledge handbook of translation studies and linguistics*. Routledge.

Moreno Sorona, L. (2018).*A comparative study of OmegaT and Memsource:*

advantages and disadvantages of their primary features. UOC Universitat Oberta de Catalunya[En línea] http://openaccess.uoc.edu/webapps/o2/bitstream/10609/74665/3/lmorenosoTFM0118memoria.pdf [Consultado el 03/11/2018]

Mor, A. (2018). Do translation memories affect translations? Final results of TRACE project. Perspectives, 27(3), 455–476.

Pym, A. (2011). What Technology does to Translating. The International Journal of *Translation and Interpreting*, *3*(1), 1-9.

(ISSN : 2536 - 9555)

Pym, A. (2013). Translation Skill-Sets in a Machine-Translation Age. *Meta: Journal des traducteurs*. *58*(3), 487-503.

Reinke, U. (2013). State of the Art in Translation Memory Technology. *Translation:*

Computation, Corpora, Cognition: Language Technologies for a Multilingual Europe,3(1), 27-48

Šanca, Filip (2018). *The Use of CAT Tools in University Translation Courses: A Case*

Study based on Teaching with Memsource. Unpublished Master's thesis. Charles University, Prague.

Teixeira, C. and O'Brien, S. (2017). Investigating the cognitive ergonomic aspects of translation tools in a workplace setting. *Translation Spaces, 6*(1), 79-103.

Vela, M., Pal, S., Zampieri, M., Naskar, S. K., and van Genabith, J. (2019). Improving CAT tools in the translation workflow: New approaches and evaluation. *In Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks, pages 8–15*.