

## Post-editing in translation: experiences and development

<sup>1</sup>Shahab Ahmad Almaaytah

<sup>1</sup>Applied College in Abqiaq King Faisal University Assistant Professor, Human science and languages- Saudi Arabia ( KSA) - [salmaaytah@kfu.edu.sa](mailto:salmaaytah@kfu.edu.sa)

### Abstract

Translation has become a very important field as globalization has made it easier for people to have access to texts on various subjects written in languages they are not familiar with. In order to meet this demand, a number of Translation Institutions have introduced a technology that combines Translation Memories and Machine Translation (MT) along with human translators in post editing , present paper investigates the skill sets required for post editors and also the possibility of developing a course to teach the basic skill sets at the University level so that a professional course can be designed for students who are getting trained as Translators and a program that includes both the theoretical and practical aspects of the course.

**Keywords:** Translation, Globalization, Translation Institutions, Technology, Translation Memories, MT, Program.

### INTRODUCTION

In the field of Translation, the conversion of a source file into a target file has seen a phenomenal change in the last several decades and this is due to the high level of automation in translation<sup>1</sup>. Post-editing is a very important part of the translation cycle, a concept that was in existence from the mid-1950 when machine translation was not producing perfect output which had to be edited manually. Post editing is a process that corrects the output to make the content better than the output the MT software has produced. This process is similar to the revision of translations done by human translators. The ‘Post Editor’ edits the content and modifies the already translated text which may have some errors such as the wrong punctuation marks, inappropriate annotations, incorrect meanings or numerals; all errors are thoroughly checked and appropriate changes are done by the post editor<sup>2</sup>. Most of the MT software give the same error in the output for the same input<sup>3</sup> and if a software is used only for getting a gist of the text then the post

editor’s job is to make appropriate corrections where necessary. The Post Editor must have the same level of competency a translator has<sup>4</sup>. A number of studies have been conducted to give a definition for the process involved in post-editing<sup>5</sup>

In a study by Aikawa et al 2007, shows the relationship of controlled language, machine translation and post editing. This paper proves that the effort put in by the post editors can be reduced if controlled language is used as the quality of MT will improve and so the overall productivity of post editing will be better. This study shows that some of the rules of CL if implemented can have a very good effect on the MT in various languages and so the difficulty in post editing becomes less. The CL rules applied in this study are limited to this specific research and so it cannot be assumed that for all MT it can be of use.

MT’s quality has improved over the decades but the quality available currently needs to be improved further in order to publish the outputs

except in cases where controlled languages are used with MT systems that are dedicated only for a few specific domains<sup>6</sup>. In majority of translation work the machine translation is considered as a raw set of data to be edited by human translators. Studies show the demand for the use of post-editing tools has grown and continues to grow<sup>7</sup>. Many workshops, publications and entries in journals have echoed the interest that translators show in post-editing MT.

The translators' productivity has also improved over the years due to the better quality of MT available as compared to manual translation<sup>8</sup>. When MT output is of poor quality it does not help in improving the translators' productivity and so the effort put in by the PE are expected to be within acceptable range which is measured<sup>9</sup>. The poor machine translation quality are edited by post editors and the basic reason for measuring the effort put in by post editors is to enable the companies to quote the correct price for the task done by the post editors<sup>10</sup>.

The usage of MT and post editing varies from language pair to language pair and similarly the rate for effort in post editing also varies based on the language pairs<sup>11</sup>. Languages that do not come under the same family are difficult to be translated into the target language as they have rules, scripts and grammar that are not present in the source language. For example English and Arabic do not come under the same family and so translation is difficult whether it is English to Arabic or vice versa<sup>12</sup>.

## Literature review

Prof. Sharon O'Brien, SALIS, Dublin City University, Glasnevin, Dublin 9 Ireland (2002) focused in his article entitled 'Teaching Post-editing: A Proposal for Course Content' on translation and its need mainly. As he said there is a growing demand for translation, to meet this demand many translation companies are introducing a hybrid technology solution combining translation memory and machine translation. However, few trainee translators

receive training in machine translation post editing. The author also emphasised on why teach post-editing, growing demand, Post-editing skills, developed gradually, teaching post-editing means translators will embrace MT, Who is the Target Audience, Translation and post-editing objectives. In totality if we say, the author trying to say is, current industry trends seem to suggest that machine translation will form an increasing part of the technology solutions put in place to meet the growing demand for translation. If this turns out to be true, a growing number of translators will have to deal with machine translation output. Let the educators of translators prepare future generations for this by teaching students about machine translation and post-editing. This paper outlines the skill sets required and proposes course content and structure. It is seen evidence that post-editing is not the same as translation or traditional revision. In fact, some of the demands of post-editing are contrary to the skills and objectives of translators and probably represent one of the reasons why MT implementation has failed in the past. Nevertheless, Mc Elhaney and Vasconcellos (ibid:142) believe that there are strong arguments in favour of training translators as post-editors. They argue that a translator is best able to identify linguistic errors, has a fund of knowledge about the cross-language transfer of concepts, and has the technical resources at their disposal to work efficiently. The conclusion, then, is that translators should be trained as post-editors. However, this type of training should be optional rather than compulsory. Qualification for a module on post-editing should be made dependent on the strengths and personality of each student, if possible.

Dr. SUN Dongyun, Lecturer, College of Foreign Languages and Literatures, Fudan University, Shanghai, China, in his article (ISSN 1712-8056[Print] ISSN 1923-6697[Online], Canadian Social Science Vol. 13, No. 7, 2017, pp. 1-5 DOI:10.3968/9698) entitled 'Application of Post-Editing in Foreign Language Teaching: Problems and Challenges' focused on Post-editing (PE) and Machine Translation (MT). He said PE is a concept that

is closely related to machine translation (MT). It “involves correcting the translation output generated by the machine translation system, a task performed by the human editor or translator in order to bring the text to a certain pre-determined standard in terms of language style and appropriate use” (Quah, 2006, p.11). Since fully-automatic high-quality machine translation (FAHQMT) is not feasible at present, post-editing MT outputs to a usable degree is a viable solution to the commercialization of MT systems. A distinction should be made between human-aided/ assisted machine translation (HAMT) machine-aided/ assisted human translation (MAHT) (Bowker, 2002, p.4) with respect to discussions on translation technology. In the former model, MT plays a major role while human translators perform pre-editing or post-editing to make results of usable. In the latter case, human translators play a major role while computers facilitate the process with translation memory, terminology management, word-processing, dictionary lookup and referencing, etc., with MT excluded. The latter is traditionally regarded as the domain of computer-aided translation (CAT) (e.g. Bowker, 2002) but is now often expanded to include MT post-editing.

Vasconcellos and Bostad (1992, p.68) believes that PE as a specialized skill should seek to retain the MT outputs and does not make modifications unless necessary. Therefore post-editing MT outputs is technically a feasible way of saving labor costs and improving efficiency. Its significance can be described in the following aspects: First, if PE is more labor-efficient than human translation, it is a cost efficient solution to commercial translation. Koehn (2009, p.23) believes that PE is cost-efficient if “the efforts of post-editing is less than the effort of translation from scratch” and that “[i]f the machine translation system reliably brings the meaning across, the post-editor does not need to know the foreign input language.” Though such an idealized scenario is unattainable at present, one can reasonably imagine the expediency of post-editing MT outputs by monolingual speakers in

lieu of the minority language professionals that are hard to find. In fact, with the development of data-driven MT such as statistical machine translation and neural machine translation complemented by domain specific data, the quality of MT outputs has improved significantly over the past decade. Garcia (2010, 2011) conducted researches of post-editing uncustomized Google MT engine outputs by professional translators and translation learners, and results showed that PE yields comparable or even superior performance in both time and quality to human translations in both cases. Second, PE is often used in benchmarking the quality of MT outputs. Despite the prevalence of automatic scoring algorithms represented by BLEU and NIST, many scholars realized the inherent defects of these algorithms as they are often a far cry from human intuition. In view of this, the Human Translation Error Rate (HTER) scheme was incorporated in the development of DARPA MT system. Here HTER is defined as “the number of editing steps divided by the number of words in the acceptable translation,” for “a human annotator has to find the minimum number of insertions, deletions, substitutions, and shifts to convert the system output into an acceptable translation.” (Koehn, 2009, p.238) Of course, the number of steps alone cannot fully reflect the actual level of difficulty of PE, and thus the time spent on human PE should also be taken into consideration.

Dr. Olga Witczak Faculty of English, Adam Mickiewicz University Poznań, Poland, in her article (Journal of Translator Education and Translation Studies, (1)1, pp. 35-55) entitled ‘Incorporating post-editing into a computer-assisted translation course. A study of student attitudes’ focused on

Translator competence in the age of technology is invariably intertwined with the ability to use digital translation aids. Thus, the work of a translator is inevitably, as Pym (2011: 4) put it, “(...) determined by internet searches, glossaries, spell checkers, grammar checkers, translation memory and machine translation databases, and anything else resembling a communication technology.” While digital

translation aids have been used by translators for some time, machine translation (henceforth MT) is still rather notorious among the translating practitioners. MT is understood in this paper as a translation produced by a computer (Oxford English Dictionary, 2016). Regardless of the widespread bias towards MT, systems generating automatic translations have been developed and researched over the years and the quality of the output is being constantly improved (Doherty, 2016). Furthermore, while the ALPAC objective of FAHQMT (Hutchins, 2007: 5), i.e. fully automatic high quality machine translation, is yet to be achieved, MT can be successfully utilised to aid human translators in a process called post-editing (henceforth PE). This process combines both the automated element (raw MT output) and the active human translator, who is the decision making agent. PE can therefore comprise “tidying up the raw output, correcting mistakes, revising entire, or, in the worst case, retranslating entire sections” (Somers, 2001: 138). This process differs substantially from translation from scratch (Carl, 2013). Modern Computer-Assisted Translation (henceforth CAT) tools integrate MT into a translation environment, thus combining PE with working with translation memory (henceforth TM) and other translation aids.

### Definitions

Veale and Way (1997) state that the task of the post editors can be made easier if the machine translation software can lessen the number of errors while translating the source text into target text. A comprehensive definition for Post-Editing has been given in the Draft of European Standard for Translation Services (2004) as “examination and correction of the text resulting from an automatic or semi-automatic machine system (machine translation, translation memory) to ensure that it complies with the natural laws of grammar, punctuation, spelling, and meaning, etc” Pym (2011b) defines post-editing as the “term for the process of making corrections or amendments to automatically generated text, notably machine-translation output.” In this paper Pym mentions that post editing is the opposite of pre-editing. According to Pym only in

automated translation situations can the two terms be used?

### Types of Post-Editing

Laurean (1984) declares that there are two types of post editing namely the conventional post editing and the rapid post editing. In the first type the contents of the Target text must be as similar to the source as possible and in the latter type the style is not as important as the language in the Target text. Temizöz (2013) states the rapid post editing as “light post editing or minimal post editing”. The main condition in this category is that content of the target text should be comprehensible for the reader. According to Temizöz (2013) this category is used in translating emails, documents that have short life or translation of messages that are to be read by very close groups of people. The time spent on such translation is less and so the light post editing or rapid post editing is executed and the output not stored in the memory permanently. The rapid post editing is most suitable in situations where there is an urgency to translate the text that need not be published, or the customer is ready to take the risk of having a low quality translation (Senez, 1998). Tremos (2013) states that in the case of full post editing the intervention of the human translators and post editors are mandatory as only then the output will have a very good quality. The target text can then be stored in MT or TM for future reference or use by translators. This process helps in reducing the time spent on translation and also increases the accuracy of the translation.

### Post Editing Categories

#### Inbound translation

Allen (2003) had clearly distinguished the Post Editing into two types namely inbound and outbound Post Editing. The Inbound post editing consisted of gisting done by the MT and any errors that is based on wrong grammar is acceptable if the Target Text can be read by the user and the meaning is not distorted. In the Inbound post editing there is another distinction

made by Allen (2003) called the Rapid Post Editing (RPE) which includes a minimal editing in which the obvious errors are eliminated. This makes the output comprehensible for the reader but here the style of the text is not changed. This kind of Post Editing was accepted by the European Commission's Post Editing services (Svěrák, 2015).

#### Outbound Translation

In the case of Outbound translation there are three categories (Martinez, 2003) namely minimal post editing, no post editing and full post editing. In the first type the draft of the text is edited with minimal changes as in the case of nay manuals or technical documents so that the final text for the reader is comprehensible. This type is used only for internal purposes of disseminating information in large corporate. The second type which does not have any post editing done on the text is similar to the original version of outputs the MT system of the 1950s and 1960s. In very few situations this can be used like in the case of weather forecasts (Svěrák, 2015).

From the above given categories it is clear that in two areas of Post editing human intervention is not required namely the gisting which is an Inbound post editing category and in the 100% MT in Outbound post editing.

#### Skills required for a Post Editor

Rico Pérez and Torrejón (2012) have identified language skill as one of the most essential skills required for a translator who is working as a post editor. In this study it has been found that among the several skills required for a Post Editor linguistic skills are very important which includes communicative skills as well as textual skills at least in two languages. Apart from the Post editor must also possess sufficient subject knowledge. The study clearly suggests the importance of language as the main requirement for a Post editor.

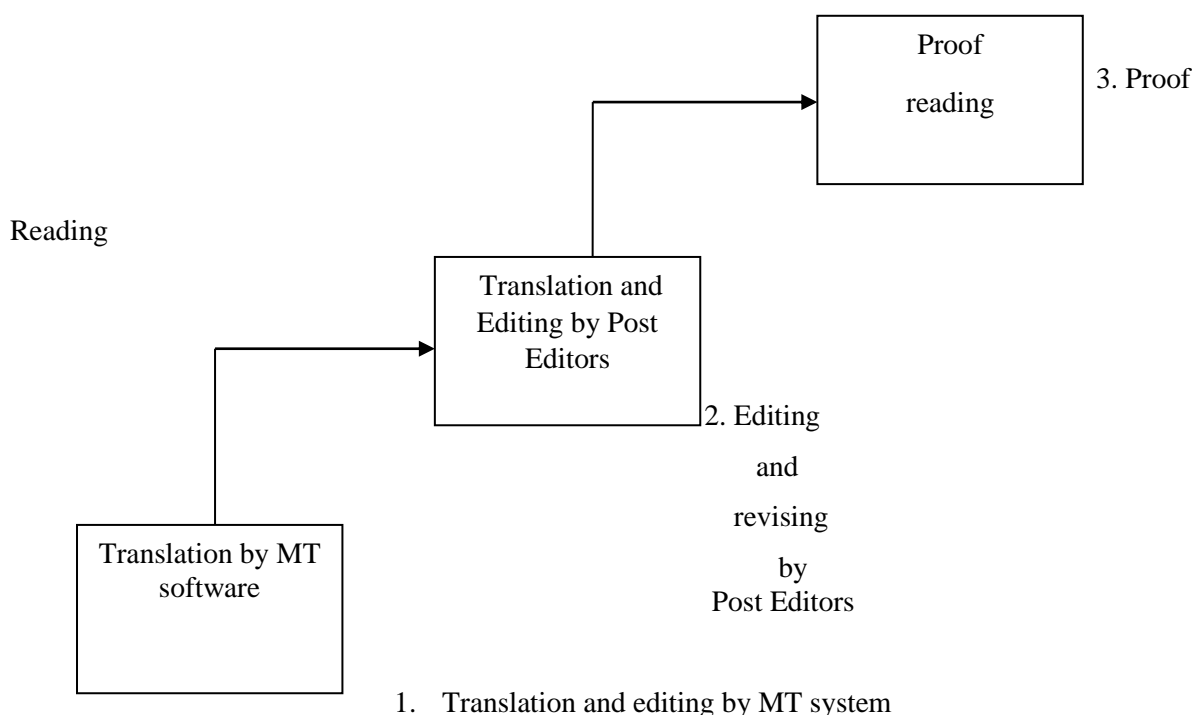
Pym's (2013) study observes that the concept of source and target language concept has been replaced by the "start text" which are

texts that are available in the translation memories, technical terms, text translated by machine translation software and glossaries which need to be translated into a specific language as per the requirements of the end user. For this the skills suggested are divided into three categories namely learning to learn, learning to trust and mistrust data, and learning to revise. Learning to learn includes the capacity of the translator to quickly grasp from online resources, estimate how suitable a given tool is for the task on hand, learning from colleagues the solution for a current problem and have skills to assess the process of the task using the tool. Learning to trust or mistrust data (Pym, 2013) signifies the capacity of the translator to either use or not the available data based on the individual discretion of the translator. Finally the category that suggests learning to revise includes correction errors such as punctuation marks, style and also learning to review work with translation professionals and subject experts.

In the language industry machine translation is a common feature that translators come across often. Machine translation is part of the curriculum in many programs that teach translation as there is a very big demand for it. Mellinger (2017) suggests that to become a professional post editor machine translation must be part of the curriculum in many courses and should not be isolated in a single translation program or course alone. The study (Mellinger, 2017) also shows that the students must be trained for the various needs of the market such as understanding, terminologies, post editing and translation practice.

The translators' expertise in more than two languages is developed due to the experience they gain from their translation work. However with the knowledge of technology the translators are better equipped to handle the task given to them.

### Workflow in Post editing



The workflow in Post Editing consists of translation by the MT software, editing the raw machine translated text by human translators or editors, and in the last step the proof reading is a task done by a human editor (Doherty and Gaspari, 2013). In this system the Translation Memory (TM) and Machine Translation (MT) are not the same but are complementary to each other and can be used to speed up the workflow. Translators and editors help in the management of the MT and they decide on the right choice of MT software. The proof reading is done by a different individual who is the Editor. It is easier to identify the errors of the MT while human errors are not very easily identifiable and so it is more difficult to set it right. So an experienced proof reader must be given the task for editing the material that is translated and edited by human translators.

#### Teaching modules for Post Editing

Some of the key areas that should be taught in a Post Editing course are MT, Editing skills, Controlled Language, basic programming skills, Text linguistic skills.

#### MT (Machine Translation)

An overview of the History and development of MT has to be part of the course as this background helps the students to understand the basics of MT. MT software should be introduced for the students to have practical knowledge and experience of a Machine translator. According to O'Brien (2002) "Knowledge of MT technology in general would go a long way towards helping the post-editor understand what is going on in the so-called "black-box" and why certain errors occur consistently."

Koponen (2015) states the importance of acquiring the knowledge of machine translation as MT is useful in various fields. This study proposes the planning and development of a course in post editing where the study of MT is a very important and significant module.

#### Editing Skills

Any Post Editing course must train students in editing skills. One of the basic skills a Post Editor must have is word processing skill. The

course must also help the aspirants to increase the speed of editing by providing practical sessions that help in increasing the speed of editing. Some of the editing skills that can be taught are

- a. Ability to make the corrections directly on screen
- b. Ability to understand the limitations of the MT that the student is trained on
- c. Reasons for the errors made by the MT used in the course
- d. Repeated sessions for improving the decision making capacity of the student
- e. Basic theoretical knowledge of the technology used in the MT and its advantages compared to the manual human translation.

This module consists of practical sessions mainly while a small amount of time can be dedicated for the theoretical knowledge.

#### Controlled language

In the translation field 'controlled language' is used for translating documents that are technical in nature. Here the language used is less complex as the grammar and style are not given importance. If the source text is input in a controlled language the translation is not difficult. If the source text has small simple sentences then translation is made very easy. If the sentences need repetition then it should be repeated which makes more than one simple sentence. If the simple sentences have simple grammar this helps the user to translate this faster. Sentences in active form using only noun and not any pronoun can be easily translated.

#### Basic programming skills

In regard to the basic programming skill that are of so much importance for post-editors, Vasconcellos (1986a:136) stresses the use of macros which is regarded as an essential skill for post-editors. According to Vasconcellos, a post-editor is an ideal candidate for writing macros to automatically clean-up texts since s/he has extensive experience of commonly occurring errors. These macros are considered

to be the first step towards the concept of an automatic post-editing tool, as suggested by Ryan (1988), Knight and Chander (1994), Allen and Hogan (2000). If the post-editor is familiar with the basic programming skill, he/she will have the ability of developing his or her own programme for automatically detecting and correcting various errors for specific language pairs, MT systems and text types.

#### Text linguistic skills

Vasconcellos (1986b) stresses the importance of knowledge of theme and rheme and other language-specific text type norms for post editing. Being familiar with text linguistic skills is of so much importance for post editors. This knowledge has to be applied for post-editing and programming macros and automatic post-editing modules.

#### Proposed outline for a course module in post-editing

In the previous section, some of the key areas that should be taught in a Post Editing course have been highlighted. The key areas include MT, Editing skills, Controlled Language, basic programming skills and Advanced Text Linguistics.

In this section, outlines of the course which addresses the key areas will be proposed.

The module in post-editing is classified into two, with a special focus on the theoretical part in the first half and a focus on the practical part in the second half.

#### -The Theoretical part

The theoretical part will include the following points:

##### 1. Introduction to Machine Translation

The Introduction to Machine Translation should cover:

- a) The history of MT
- b) MT system types
- c) Description of commercial MT systems

- d) Evaluation methodologies
- e) The state of the art, including integration with translation memory tools,.
- f) Future prospects.

## 2. Introduction to post-editing skills

This section covers the following:

- a) The concept of post-editing
- b) The need for post-editing
- c) How is post editing different from translation and revision,
- d) The various levels of post-editing,
- e) Identifying the user requirements
- f) The technology which is needed for post editing,
- g) Classifying typical post-editing errors.

## 3. Introduction to Controlled Language

This section covers the following:

- a) A history of Controlled Language,
- b) A description of the different CL tools,
- c) Evaluation methodologies for CL tools
- d) current state of the art
- e) Integration with authoring and MT tools,
- f) Future prospects

## 4. Basic Programming Skills

This section introduces the following:

- a) The basics of programming
- b) Macro programming
- c) Programming language suitable for Natural Language Processing, for example Perl.

## 5. Text linguistic skills

This part covers:

- a) The basic linguistic skills
- b) The standards of textuality

## c) The text type classification

- d) The use of corpus linguistics and corpus analysis tools for analysing text types.

Practical part:

Regarding the practical part of the key areas, the following should be part and parcel of the practical component:

### 1. Introduction to Machine Translation

In this section, the students should be trained to do the following:

- a) Practise the Commercially-available MT systems .T
- b) Submit different texts for translation to the MT system
- c) Analyse and compare the results under different system settings
- d) Explore the pros and cons of a MT system's integration with a translation memory tool.

### 2. Introduction to post-editing skills

It is assumed that post-editing is a practical skill and teaching this important skill aims to let the students acquire the "comfort" factor Vasconcellos talks about before being recruited, practical experience of post-editing is considered to be a major component in the course. According to Vasconcellos (1986a:145), a post-editor at PAHO (the Pan-American Health Organisation) post-edits 100, 000 words, or almost one full working month, before that level of comfort is reached. While it may not be possible for a student to attain this goal, especially considering the workload from the theoretical component of a programme.

In this section, the student should be trained to do the following:

- a) Practising post-editing both within and outside course hours.
- b) Post editing of various text types from many MT systems
- c) Post-editing into multiple target languages.



d) Practising the different “levels” of post-editing.

### 3. Introduction to Controlled Language

Practical experience of controlled language tools should be gained according to the following guidelines:

- a) Checking and editing texts in the source language using a CL tool
- b) Submitting the controlled and uncontrolled texts to a number of MT systems.
- c) Post-editing of both versions and showing the pros and cons of controlled authoring for machine translation.

### 4. Basic Programming Skills

In this section, the students should acquire practical programming skills by doing the following:

- a) Writing macros to automatically apply common changes in target texts.
- b) Applying the programming language skills learned in the theoretical component of the course by designing a rudimentary automatic post-editing application.

### 5. Text linguistic skills

The students should gain practical experience on:

- a) corpus analysis by compiling parallel corpora, tagging them, and analysing them for specific text linguistic features such as theme/rheme structure, voice, tense, cohesive ties, etc. using corpus analysis software such as Wordsmith tools.

## Conclusion

Recently, translation has witnessed significant and dramatic changes brought about by the advent of new technologies such as machine translation. The output of the MT has to be post edited by human beings and this process is called post editing .In this process, humans post edit machine-generated translation to come up with an acceptable final product. The post

editors should be trained to tackle different terms such as translation, post editing and its types, post editing categories and the skill sets required for post editors .It is highly recommended that designing a post editing course is essential in order to be taught at the university level to produce professional post editors. For this purpose, this paper outlined a program that includes both the theoretical and practical aspects of the course.

This work was supported through the Annual Funding track by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Project No. AN000114'

## Reference

- [1] Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and Translation: A Translators Guide*. Benjamins Translation Library, 35. Amsterdam: John Benjamins.
- [2] Arenas, A. G. (2013). What do professional translators think about post-editing. *The Journal of Specialized Translation*, (19).
- [3] De Sousa, S. C., Aziz, W., & Specia, L. (2011, September). Assessing the Post-Editing Effort for Automatic and Semi-Automatic Translations of DVD Subtitles. In *RANLP* (pp. 97-103).
- [4] Farghaly, A., & Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 8(4), 14.
- [5] Groves, D., & Schmidtke, D. (2009). Identification and analysis of post-editing patterns for MT. In *Proceedings of MT Summit* (Vol. 12, pp. 429-436).
- [6] Koponen, M. (2015, November). How to teach machine translation post-editing? Experiences from a post-editing course. In *4th Workshop on Post-Editing Technology and Practice (WPTP4)* (p. 2).
- [7] Koponen, M. (2016). Is machine translation post-editing worth the effort? a survey of research into post-editing and effort. *The Journal of Specialised Translation*, 25, 131-148.

- [8] Laurian, A. M. (1984, July). Machine translation: What type of post-editing on what type of documents for what type of users. In *Proceedings of the 10th international conference on Computational linguistics* (pp. 236-238). Association for Computational Linguistics.
- [9] O'Brien, S. (2002). Teaching post-editing: a proposal for course content. In *6th EAMT Workshop Teaching Machine Translation* (pp. 99-106).
- [10] O'Brien, S. (2005). Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability. *ResearchGate*, 19(1):37-58 DOI:10.1007/s10590-005-2467-1-DBLP.
- [11] Plitt, Mirko and François Masselot (2010). "A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context." *The Prague Bulletin of Mathematical Linguistics* 93, 7–16.
- [12] Schütz, J. (2008, October). Artificial cognitive MT post-editing intelligence. In *AMTA-2008. MT at work: Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas, Waikiki, Hawai'i* (pp. 448-453).
- [13] Temizöz, Ö. (2013). Postediting machine translation output and its revision: Subject-Matter Experts versus Professional Translators.