Teaching Computer-assisted Translation: A Proposal for Course Content

M. Brkic Bakaric
Department of Informatics, University of Rijeka, Rijeka, Croatia
mbrkic@uniri.hr

Abstract – With wider adoption of machine translation and its integration in the standard translation workflow, the job of translator developed into various forms – post-editors, revisers, translators, proof-readers, etc. In this paper we argue for the inclusion of the post-editing, pre-editing, and machine translation content into the course on computer-assisted translation and give proposal for the course content. We also argue for the inclusion of added value services such as subtitling and localization. The course should be mandatory for all the graduate students studying a language. The content is influenced by the experiences gained from teaching a similar elective graduate-level course.

Keywords – computer-assisted translation course; course content; translators’ technical competence; added value services

I. INTRODUCTION

It is clearly obvious that pen-and-paper translation has come to an end [1]. As the author in [2] puts it, everyone agrees that there is need to train translators to meet the existing demand and to face rising volumes in the future. Neither those who graduate from a translation study, nor those who are trained “at the job” are necessarily good translators. It turns out that the major concern is how to train well in order to get good translators.

Users may be prevented from using machine translation (MT) due to a lack of awareness of the current capabilities of MT systems and lack of technological know-how [3]. Insights into the practice can be gained from the survey in [4].

According to the standard ISO 17100 [5] translators have at least six competences. Technical competence relates to the knowledge, abilities, and skills required to perform the technical tasks in the translation process by employing technical resources including the tools and IT systems that support the whole translation process. In addition to this, there is also research competence which requires experience in the use of research tools and the ability to develop suitable strategies for the efficient use of information sources available.

An elective course ICT for translators started to be conducted at the University of Rijeka as of 2016/2017. However, the course was offered only to students of single or double major Informatics. In such a way only students of a language in combination with Informatics could be enrolled to the course. There are four language departments at the University of Rijeka – Croatian, English, German, and Italian. Due to a positive feedback, in September 2017 a one-day workshop was organized through which the most representative processes of the professional translation workflow tried to be approached. The course is as of 2017/2018 open to students of all university constituents, not only language graduates, in line with the opinion in [2]. The experience gained is drawn and a proposal for the obligatory course of a graduate study in Translation Studies is given.

Section II introduces basic concepts in the field of MT and computer assisted translation (CAT). Related work is presented in section III. Section IV gives the course outline. Evaluation is provided in section V, and a short conclusion is presented in section VI.

II. BASIC CONCEPTS

Standards can be seen as minimum requirements which must be accounted for by training institutions if they want to ensure market-oriented training [6]. The standard ISO 17100 sees the training of translation service providers as a life-long learning process. As the author in [6] points out, successful translators continuously work on their self-improvement.

According to the standard, the translators should have:

- a degree of the first academic level in translation or linguistic studies or language studies which include translation training, or
- a degree in the domain and two years of full-time professional experience in translating, or
- at least 5 years of full-time professional experience as a translator.

From the definition of the standard it is obvious that translator tasks are so diverse that it is not possible to have a unified educational framework. A non-exhaustive list of added value services is included in the appendix of the standard.

Since there might be some confusion about the terms and different roles that translators take in the translation workflow, a definition of the basic roles is given in Fig. 1.

This research has been supported under the Grant No. 17.14.2.2.01 of the University of Rijeka and Project No. uniri-drustv-18-122.
The author in [7] makes a distinction between PE on one hand, and translating and proofreading on the other. While in PE it is of crucial importance to check every sentence against its source counterpart in order to identify “tricky” MT mistakes, proofreading checks meaning, comprehensibility and idiomaticity.

III. RELATED WORK

The significance that software expertise has for translators is underlined in numerous research papers [1, 7, 8]. Not only should the basic version of a translator’s workstation be taught, but also tools for subtitling and localization. A basic introduction to MT is also important [9]. A holistic approach to the integration of MT into translator training programs that empowers and not marginalizes translators, along with a non-technical overview of statistical MT is presented in [10]. The design and evaluation of the related syllabus is presented in [11].

The aim, as the author in [2] puts it, should not be to train students for a narrowly defined market, but to empower them to apply for a wide range of positions in translation. IT as an ever-growing set of tools should ensure mastery of all technology available in a state-of-the-art translator’s workstation, the ability to analyze and appraise any software package, and the ability to train colleagues in the use of such tools [2]. The author in [2] goes even further and says that among other things, translators must master all the translation-localization (and revision) techniques, methods and procedures and be totally proficient in the use of all the equipment and software they are likely to come across in their professional environments. Even back in 2007 the market was eager to take on people trained in special skills areas such as software, web site, and videogame localization, post-editing, etc. The author in [12] defines professional translation even as one form of human-computer interaction.

The need for post-editor training has been emphasized already in [13]. Training post-editors is further highlighted as necessary in [8, 14, 15]. PE skills should be taught because of the increasing demand for translation and for faster production times, and because they differ from translation skills. Teaching these skills could improve the uptake of MT technology by improving translators' perceptions of MT and its capabilities [8]. Since PE skills are developed gradually, companies implementing MT would benefit if translation graduates were comfortable with PE. Hence, graduates would be given better employment opportunities. A strong argumented case for teaching PE is given in [8]. Knowledge of MT, understanding the history, and understanding the future prospects of MT are important in gaining appreciation for MT and understanding its limitations and possible improvements. Post-editors should be taught how to code dictionaries and manage terminology databases (TB). They should be acquainted with multiple terminology management tools and exchange formats. Post-editors should also have the knowledge of controlled language and controlled authoring tools in order to play the role of an intermediate editor who authors source texts with the aim of improving MT output. The objective of the work in [15] is to model PE behavior in order to design specifications for computer-aided PE and a training program for new post-editors. PE serves multiple purposes. It perfectly fits the scenario of MT evaluation because it gives concrete measures of translation quality like time and effort spent in PE, and it reveals the gap between MT output and human translations [16]. Students can be organized in groups in order to devise guidelines for different types of PE themselves, bearing in mind how long a correction will take and whether a correction is necessary. They should do all kinds of PE on the same text in order to see the difference more clearly. Students should work from the foreign language to their native in order to suppress the fear of PE making them learn incorrect models. The triad of controlled languages, MT
systems, and post-editing processes, coupled with other CAT tools such as translation memories (TMs), is called controlled translation by the authors in [17]. A machine translation and post-editing course which combines this triad is presented in [18].

Specialization by job and technology and domain was a marked major trend fifteen years ago meaning that translators identified themselves primarily as localizers, or subtitlers, or revisers, etc. The author in [2] advocates that all students give priority to anything that goes by the name of localization since no one can predict the combination that will prevail for any individual student.

The potential of the online platform Crowdin for the education of future translators is highlighted in [19]. It is used for translating of websites, mobile and desktop applications, etc. Since it supports crowdsourcing, it can be used even without the need to create special content.

The author in [2] emphasizes that all students have to have the necessary language skills and competences required for translator training courses. We take a bit different approach and treat this course as independent of the language skills. We aim to develop the necessary IT skills which can be blended with language skills from the very beginning, gradually during the course, or afterwards.

IV. BUILDING BLOCKS

The aim of the course is not to master all the tools with respect to the time dimension, but only to familiarize students with types of tools out on the market. Similarly to [20], we believe that basic competence can be built on once the students begin to work as professional translators. The course is divided so that each theoretical component is followed by a practical one. Practical labs take place in computer labs with specialized software.

In the course proposal a special attention is paid so that as many of the non-exhaustive list of added value services are included. A selected subset of these services consists of pre and post-editing, localization, revision, subtitling, translation memory alignment and alignment of bilingual parallel texts, and, lastly, terminology management. These services increase employability of translators and engagement of language service providers (LSP).

The course is envisaged as a hands-on experience and all the focus is put on practical work. As emphasized in [4], effective teaching in these fields requires hands-on practice in computer labs and the use of dedicated software. During class, students are guided through exercises followed up by real-world homework assignments. The class is divided into eight assignments and a final exam in the form of a project (Fig. 2). The Moodle platform is used for distributing materials and collecting assignments. For their final exam students have to pick or propose a topic which can be an extension of a class lab or a new topic, such as video game localization, tutorial for a software, comparison of different tools, etc., and they have to cover its theoretical and practical aspects. Examples of projects suggested and conducted by students are given in Table 1.

Since most of the tools and services are free and/or open-source, many assignments can be done at home in the form of distance learning, similarly as in [11]. This is important due to the sudden acceleration of technological changes affecting the profession which raises the need for continuous education and training.

Besides the building blocks described in the remainder of this section, one session should be reserved for a practicing professional as a guest-lecturer and a significant amount of time should be devoted to the question-answering section.

A. Corpus compilation and analysis

This section of the course is concerned with getting the best out of electronic corpora. Corpus use for learning to translate and learning corpus use to translate are two complementary approaches to the use of corpora in translation. The latter is concerned with helping students to become autonomous users of corpora as part of their translation competence [21]. We take, as [22] puts it, more macroscopic perspective in that students themselves have to compile the corpora before they can apply them to

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE + automatic evaluation</td>
<td>PE is done using PET tool on randomly selected sentences of the Workshop on MT (WMT) testset translated by Google Translate (GT) from English into Croatian and from Italian to Croatian. Correlation of the editing time and a chosen automatic metric result per sentence is calculated.</td>
</tr>
<tr>
<td>Subtitling</td>
<td>Karaoke in language learning</td>
</tr>
<tr>
<td>Error analysis + automatic evaluation</td>
<td>Three text types are translated from German into Croatian by GT. Automatic metrics chrF and BLEU are calculated, error analysis is performed, correlations are investigated.</td>
</tr>
<tr>
<td>Subtitling</td>
<td>Using text-to-speech captioning and MT subtitling - the trade-off between quality and time</td>
</tr>
</tbody>
</table>
solve any translation problems. Students are first introduced to the characteristics of conventional search engines and familiarized with the tools and strategies for querying the Internet as a macro-corpus. Afterwards, they are given guidance in using special concordance software for compiling and querying monolingual and bilingual corpora.

B. Machine translation and evaluation

Since MT is nowadays integrated in the majority of CAT tools, translators and post-editors must be aware of the capabilities of MT systems. Error analysis is a means to assess translations in qualitative terms. It refers to the identification and classification of individual errors in a translated text [23]. Students are given a detailed introduction into the human and automatic evaluation of machine translation and various taxonomies and tools are presented for performing evaluation and error analysis. By comparing different automatic error metrics scores with human judgements of the quality of MT output, the students develop ability to identify strengths and weaknesses of different approaches [11]. Two paths that can be followed during the lab session are presented in [24]. Students can either evaluate the same MT service on texts from different domains or evaluate the same text translations provided by different MT services.

C. Pre-editing

In line with the research in [17] and [25], we believe that translation programs should teach pre-editing as well. The author in [26] advocates pre-editing as one of the most exciting and effective ways of learning to use MT, learning to translate, and learning the language. An overview of different pre-editing guidelines can be found in [17]. Similarly to [11, 17, 18, 21], we ask trainee translators to input a foreign language sentence into an MT system, observe the errors in the translation, edit the source text by following a basic set of instructions or their intuition, and input it again in the hope of getting better output from the MT system.

D. Post-editing

PE, as [16] puts it, allows for discussion of the most abstract principles which underlie authoring and translation processes. On the other hand, it gives students their first experience of one of the most practical skills which trainee translators are expected to acquire. Students do rapid and full post-editing on the same text in order to see the difference more clearly, and bearing in mind how long a correction will take and whether a correction is necessary. A set of example cards illustrating guidelines is provided in order to make the training proper, as in [27]. A concise overview of guidelines and an actual exercise which provides students with hands-on experience on all the key elements of the controlled translation is presented in [17]. Another interesting exercise can be post-editing without source text [18].

E. Translation memories and terminology databases

Working with translation memories might appear cumbersome at the beginning because no immediate benefits are seen. After theoretical introduction to TMs and ethical considerations related to ownership and confidentiality, the learning process is somewhat similar to the model proposed in [28]. It is divided into several phases (Fig. 3). Similarly to [8], in this session we propose discussing the strengths and weaknesses of terminology management tools, dictionary coding for MT, and terminology exchange between tools using terminology exchange standards. Students also learn how to enable or disable machine translation, if the tool offers such support. The same applies to the publicly available translation memories. In order to complete the assignment, students have to upload the final translation, as well as the exported TM and TB.

Although there are free and open-source tools and services available, it is important that students get familiarized with at least one commercially available tool. We obtained a free academic license for one such tool. Since students do not know what tools they will have to use in their future work, the aim is to get them acquainted with as many tools as possible. The more tools and services they work with, the more they are accustomed to working with TMs in general, as the underlying processes do not differ much. The resources created with the first tool are migrated to a new tool as they move along, similarly to [20].

In order to experience different roles, when the tool or service allows it, students are asked to take up the role of the project manager and assign translations to translators.

We devote four sessions for this central part of the course. After each session, students are asked to reflect on the advantages and disadvantages of the introduced tool and to compare it to similar tools they have already worked with.

F. Subtitling

The author in [29] advocates subtitling in translators’ training with the aim to fine-tune translation skills, develop critical thinking, learn to take distance from word per word translation, and concentrate on semantic units. There is a number of semi-professional tools on the Internet which can be used in classroom in order to enable students to meet the audiovisual translation and its particularities.
G. Localization

The author in [19] suggests using Crowdin as an introduction to the use of commercial CAT software. We, on the other hand, find appropriate introducing it only after free and commercial CAT tools have been introduced. We use it to localize applications developed by the students of Informatics. Furthermore, a cooperation has been initiated with the Raspberry Pi Foundation. The aim of this UK-based charity is to put the power of computing and digital making into the hands of people all over the world. These sorts of materials can also serve the purpose of introducing special vocabulary in the field of computer science, which is by far the fastest developing field, especially in terms of language.

V. Evaluation

Since the course needs to be up-to-date, it will be continuously updated and improved. This is emphasized also in [11]. For that purpose we conduct different evaluations.

The workshop is evaluated with a questionnaire measuring the satisfaction with the organization of the workshop, topics covered, presentations, and labs. A general opinion on the usability of the workshop for future work is also tested. The average grades on a Likert 1-5 scale are 4.6 or above for all the questions, 5 being the dominant value per each question, with standard deviations below 0.17, and a total number of participants 10.

A questionnaire measuring the attitude of students toward MT and CAT tools is administered during the first and last week of the class, i.e. before and after taking the newly designed course. As in [24], students are oblivious to the fact that they will answer the same queries twice in order to make their responses in the second phase as spontaneous and as objective as possible. Due to the absence on a pre-test or a post-test, the test statistics consists of only 6 samples. The objective is to test multiple hypotheses:

a) A newly designed course positively affects students' attitudes toward MT.

b) A newly designed course positively affects students' attitudes toward TMs.

c) A newly designed course positively affects students' attitudes toward PE.

All three hypotheses are inspected in their general form, as well as with respect to time and quality dimensions. Although the test statistics is too small to draw any statistically significant conclusion, a change of opinion can be detected for the case of PE. It is also interesting to note that although the general attitude towards TMs is positive, as well as the attitude with respect to the time dimension, a fear for the final quality exists. It seems like the least affected attitude is the one towards MT. This can be attributed to the fact that MT is widely available and that students already use it on a daily basis.

VI. Conclusion

The paper gives a proposal of the course on computer-assisted translation. The course is envisaged as a hands-on experience and all the focus is put on practical work. Each theoretical component is followed by a practical one. During lab sessions, students are guided through exercises followed up by real-world homework assignments. The class is divided into eight assignments and a final exam. A special attention is paid so that as many of the non-exhaustive list of added value services are included in the course content.

The course has been constantly developed for three years to gain the form proposed here. Since the nature of the course does not allow it to be out-of-date, it will be
continuously updated and improved as the market dictates. In order to do this methodologically, different evaluations are conducted. However, the presence of different cohort each year provides a degree of uncertainty since the data may be skewed by inconsistencies between the cohorts, as acknowledged in [20]. Moreover, for students of humanities and social sciences technology may present a bigger challenge than to those in some other areas. In addition, providing students with materials specific to their own language pairs is not always feasible within the scope of the course. However, all the assignments require them to work from the foreign language or from English to their native.

The aim of the presented course is not to master all the tools with respect to the time dimension, but only to familiarize students with types of tools out on the market as the basic competence can be built on once the students begin to work as professional translators.

ACKNOWLEDGMENTS

This work has been fully supported by the University of Rijeka under the grant number 17.14.2.2.01 and project number uniri-drustv-18-122.

REFERENCES