Routledge
Taylor & Francis Group

Perspectives
Studies in Translation Theory and Practice

ISSN: 0907-676X (Print) 1747-6623 (Online) Journal homepage: http://www.tandfonline.com/loi/rmps20

Embracing the threat: machine translation as a solution for subtitling

Lindsay Bywood , Panayota Georgakopoulou & Thierry Etchegoyhen

To cite this article: Lindsay Bywood , Panayota Georgakopoulou & Thierry Etchegoyhen (2017) Embracing the threat: machine translation as a solution for subtitling, Perspectives, 25:3, 492-508, DOI: 10.1080/0907676X.2017.1291695

To link to this article: http://dx.doi.org/10.1080/0907676X.2017.1291695

PERSPECTIVES, 2017
VOL. 25, NO. 3, 492–508
https://doi.org/10.1080/0907676X.2017.1291695

Routledge
Taylor & Francis Group

Embracing the threat: machine translation as a solution for subtitling

Lindsay Bywood [a], Panayota Georgakopoulou[b] and Thierry Etchegoyhen[c]

[a]Department of Modern Languages and Cultures, University of Westminster, London, UK; [b]Deluxe Localization, Athens, Greece; [c]Department of Human Speech and Language Technologies, Vicomtech IK4, Donóstia-San Sebastián, Spain

ABSTRACT

Recent decades have brought significant changes in the subtitling industry, both in terms of workflow and in the context of the market for audiovisual translation (AVT). Machine translation (MT), whilst in regular use in the traditional localisation industry, has not seen a significant uptake in the subtitling arena. The *SUMAT* project, an EU-funded project which ran from 2011 to 2014, had as its aim the building and evaluation of viable MT solutions for the subtitling industry in nine bidirectional language pairs. As part of the project, a year-long large-scale evaluation of the output of the resulting MT engines was carried out by trained subtitlers. This paper reports on the impetus behind the investigation of MT for subtitling, previous work in this field, and discusses some of the results of this evaluation, in particular an attempt to measure the extent of productivity gain or loss for subtitlers using MT as opposed to working in the traditional way. The paper examines opportunities and limitations of MT as a viable option for work of this nature and makes recommendations for the training of subtitle post-editors.

ARTICLE HISTORY
Received 1 March 2016
Accepted 30 November 2016

KEYWORDS
Statistical machine translation; subtitling; audiovisual translation; *SUMAT* project; productivity

1. Introduction

To say that recent decades have brought enormous changes in the way we communicate and consume entertainment seems almost fatuous, so self-evident is this fact. A direct result of this phenomenon is the change in the status of and demand for translation. As communication between people in all corners of the world has become practically instantaneous, those same people have begun to demand to not only understand what others are trying to convey but also access the same entertainment material, at the same time, irrespective of the source language. This is a subset of a wider phenomenon, whereby people now expect to be able to obtain a rapid translation for most of the text they encounter in their daily lives; a phenomenon described by Zuckerman as 'The Polyglot Internet'. Zuckerman (2008) states: 'for the Internet to fulfil its most ambitious promises, we need to recognize translation as one of the core challenges to an open, shared and collectively governed internet'.[1]

This rising demand for multilingual content, of which a significant part is audiovisual in nature, raises important questions for the subtitling industry, which is asked to localise

CONTACT Lindsay Bywood L.Bywood@westminster.ac.uk

an increasing amount of material in constantly shrinking turnaround times, whilst finding new cost-effective ways to do so. The localisation industry at large has responded to the same issue with the emergence of cloud-based platforms, workflows involving crowd workers and the incorporation of technological solutions, such as machine translation (MT), to aid the job of translators (or replace them in the case of ephemeral textual material that only requires gist translation). It makes sense that the subtitling industry also looks at adopting these solutions. At the same time, freely available MT, with widely used systems such as Google Translate, has raised the visibility of the tool (Doherty, 2016, p. 953), even more so since Google Translate's integration into YouTube in 2008 for the automated translation of subtitles (YouTube Official Blog, 2008).[2]

The advent of MT in the subtitling process needs to be closely examined, given both its potential contribution in terms of productivity and its impact on established subtitling practices. We explore and discuss these aspects in the remainder of this article, drawing on the results of the evaluations of the usability of Machine Translation (MT) subtitle output that took place in the *SUbtitling for MAchine Translation* (*SUMAT*)[3] project. *SUMAT* was an EU-funded project which aimed to build a cloud-based service for the MT of subtitles in nine languages grouped in seven bidirectional language pairs. The project was carried out by a consortium of nine partners, four of which were subtitling companies, the staff of which spent a full year evaluating such MT output. This was the first extensive evaluation of MT output by professional subtitlers and, as such, its results are pertinent in terms of informing decisions regarding the application of MT in subtitling workflows.

2. The subtitling industry

The subtitling industry has experienced various seismic shifts in the course of the past three decades, each of which has defined it and shaped its future course. The first of these was the cable and satellite TV revolution in the late 1980s, which greatly increased the quantity of content to be subtitled for television viewers across the globe. English-language broadcasters, previously uninterested in non-English markets, could now reach thousands of new viewers with their existing content for the price of a subtitle file. This market expansion led to the founding of the forerunners of today's large international subtitling companies, as entrepreneurs realised the need for third-party companies to service this increased demand.

The next event was the introduction of the DVD in the 1990s, which not only took subtitles into the domestic household but also led to the spread of subtitling to traditional dubbing and voice-over countries, and another increase in content volumes requiring interlingual subtitling.[4] As a consequence of the rapid growth of the DVD market, a concurrent workflow change took place in the industry with the introduction of the template method, which involves producing a master subtitle file in the source audio language (template) for translation into all the target languages required (Georgakopoulou, 2006). This development made it possible to allocate work to translators with minimal training in the art of subtitling, as they only had to apply themselves to the textual content of a subtitle, and no longer needed to be well-versed in the technical needs of the profession.

The next milestone in subtitling history was the growth of high-speed broadband and the advent of Web 2.0 (Graham, 2005), which led to the proliferation of audiovisual

content online, much or all of which was suitable for subtitling. In addition, content providers have been facing ever-increasing demands to make their intellectual property (IP) available in multilingual versions more and more quickly. The trained subtitling workforce has been experiencing continued pressure as a result of these developments. On top of this, the global economic crisis has had the effect that, whilst broadcasters are continuing to expand into new markets, their revenue is remaining static or falling (House of Lords Select Committee on Communication, 2011, p. 60), leading to significant price pressure on those companies providing audiovisual translation (AVT). The need to find faster and cheaper ways of producing subtitles has never been more pressing, and MT has been proposed as a way to increase productivity and hence shorten turnaround times, while also reducing project costs and allowing trained professionals to work for the rates which the market will support.

3. Practice and tools

The subtitling process has undergone many changes since the start of large-scale commercial subtitling in the late 1980s. The subtitler's working conditions, as for most creative workers, have changed from the laborious (pen and paper) to the convenient (desktop computer or laptop and time-saving tools). However, despite these changes, the subtitling industry has not benefited from the same level of translation automation as the traditional text localisation industry. It is true that top-end subtitling software now includes many tools designed to make the subtitler's life easier, such as shot-change detectors, sound wave representation, autotime functions, and automated quality checks for subtitling-specific issues (Georgakopoulou & Bywood, 2014); however, there has been little or no uptake in the subtitling industry of other software solutions, such as translation memories (TMs), translator's workbench functionality, and MT. This is particularly surprising as the text localisation industry has been using such tools for approximately two decades now.

Improvements in the standards of MT have led to increasing acceptance and integration of MT technology into standard workflows in the wider localisation industry (University of Edinburgh, 2014), not only for technical domains, such as manuals, software documentation, and knowledge bases, but also for more creative texts, such as hotel websites (Reid, 2013) and marketing texts. Texts requiring real-time communication and involving online user activity, such as email support, live chat, Facebook posts, customer reviews (PROMT, 2012), ecommerce (Wohlsen, 2014), etc., have become great test cases for unedited MT, while agile localisation workflows today would not be possible without the implementation of MT, e.g. the localisation of Dell.com in 27 languages (Barbour, 2013). This is not true, however, of the subtitling industry, which in this respect lags significantly behind.

Given the apparent lack of interest from the subtitling industry in embracing MT technology compared with the traditional text localisation industry, it is worth exploring some of the possible reasons for this reluctance. The subtitling domain presents two aspects that render automatic translation all the more difficult. First, it is an open domain, using vocabulary from any imaginable number of domains – news, education, lifestyle, sports, law, teen slang, etc. – and the success of MT has been proven to increase the more the vocabulary and the domain are restricted (Bender, 2010, p. 6). Second, the source text (ST) in subtitling is a written representation of spoken language, with its own grammatical

specificities (McCarthy & Carter, 2001; Paterson, Caygill, & Sewell, 2012). Spoken language translation poses challenges for most language analysis tools, as they have been developed with written grammar in mind. Rule-based MT systems, for instance, were typically developed by assuming fairly standard grammar rules and face serious difficulties when confronted with spontaneous speech with, for example, elided forms, mixed interjections, and interrupted speech (Matusov, Ney, Popovic, & Zens, 2004).

Despite this, there have been a handful of attempts to introduce MT in subtitling. Initially this was with rule-based MT systems applied in captioning, which shares many common characteristics with subtitling. Popowich, Mcfetridge, Turcato, and Toole (2000) attempted to build a system called *ALTo* for MT captions of North American broadcasts from English into Spanish. The evaluation of the system output by Spanish speakers was favourable, identifying 70% of the translations produced by the system as correct or acceptable, and 41% as correct. Soon after, a patent was filed for the software *TranslateTV*,[5] which has been in use since 2003, providing automated LA Spanish live captions on the basis of their corresponding US English ones.

In Europe, the EU-funded project MUSA[6] ran between 2002 and 2004 with the aim of developing a system that would combine speech recognition, text analysis, and MT to create multilingual subtitles in English, French, and Greek (Piperidis et al., 2004, p. 205). This ambitious project aspired to automate the entire subtitling process end-to-end by integrating different technologies in order to convert English audio into transcribed text, generate English subtitles from these transcripts by condensing the text so as to abide by the spatio-temporal constraints of subtitling, and, finally, MT these subtitles into French and Greek (Piperidis, Demiros, & Prokopidis, 2005).

Another two-year European project, eTITLE[7] (2003–2004), also attempted to build a system that would integrate tools such as speech recognition for text to audio alignment purposes, text condensation, TMs, and MT in order to aid subtitlers in their work. Melero, Oliver, and Badia (2006) worked in Catalan, Spanish, English, and Czech. They did not train their own MT system with relevant data but simply resorted to freely available systems. A small productivity gain evaluation in the most difficult language of the project, Czech, demonstrated a 17% time benefit from using MT versus human translation, despite the poor performance of the MT with regard to Czech morphology (Melero, Oliver, and Badia, 2006, p. 17).

The use of TMs and MT in subtitling was also explored by O'Hagan (2003) in a preliminary study, which served as the basis for a more thorough investigation of the application of MT to the subtitling process by Armstrong, Caffrey, and Flanagan (2006). The latter was the first time example-based MT was applied in the subtitling domain; the language pairs investigated were English–German and English–Japanese. The results were low in terms of automated MT quality metrics, and the user evaluation was also limited; however, one should also note the low volume of data available to train the systems.

Data-driven approaches to MT have seen a significant rise since the turn of the century, and statistical machine translation (SMT), in which translation knowledge is built from corpora of naturally occurring language (Koehn, 2010b), has dominated the market until now. In 2008, Volk reported on the first significant commercial application of MT in subtitling, in which an SMT system was built to translate subtitles from Swedish to Danish and Norwegian (Volk, 2008). A large corpus of data was used to train the

system (5 million subtitles) and the output was found to be of such good quality that a commercial deployment of the system ensued and further language pairs were added. This paved the way for more work on the subject.

4. The *SUMAT* project

Subtitling has previously been recognised as an area that could greatly benefit from the introduction of SMT technology to increase translator productivity (Castilho Monteiro de Sousa, Aziz, & Specia, 2011; Hardmeier & Volk, 2009; Volk, 2008), partly because subtitles are considered to be ideal training material for SMT as they are short, grammatically discrete units (Volk, 2008, p. 7). The use of templates for well over a decade has meant that an increasing quantity of professionally-produced, high-quality parallel corpora of subtitles has been and continues to be generated by subtitling companies. These corpora are of great interest to MT researchers; however, most large archives of professional subtitle data are the property of subtitling companies and their clients, which make them difficult for MT researchers to access.

The *SUMAT* project addressed this issue by inviting major subtitling companies to be part of the project consortium. This 3-year long EU-funded project was a collaboration between four subtitling companies and five technical partners. The subtitling companies were Deluxe Media Europe,[8] Voice & Script International,[9] InVision Ondertiteling,[10] and Titelbild Subtitling and Translation,[11] two large multinational subtitling companies with offices around the world and two leaders in subtitling in their respective local markets. These companies took on the role of data providers and system evaluators in the project. The five technical partners were CAPITA TI[12] and TextShuttle,[13] two translation companies specialising in MT solutions for industry; ATC[14] and Vicomtech-IK4,[15] two research centres; and the University of Maribor.[16]

The project was completed in early 2014 and included the most extensive evaluation to date of SMT output for subtitling purposes by professional translators. During the course of the project, a total of over seven million parallel subtitles in seven language pairs (English into and out of German, French, Spanish, Swedish, Portuguese, and Dutch, and Serbian into and out of Slovenian) as well as 15 million monolingual subtitles were collected by the subtitling companies from their archives and prepared by the research partners to train the MT systems. The research partners built the relevant SMT engines, and the systems were systematically refined through various techniques adapted for the correction of recurrent errors. Large amounts of freely available corpora of both professional and amateur quality (approximately 110 million aligned segments in total) were also used to extend the coverage of the systems and various trials were carried out combining the professional subtitle data collected during the project with such freely available data.[17] The output generated from the various combinations was subject to comparison in terms of quality and the best systems were selected for the final evaluation (Etchegoyhen, Fishel, Jiang, & Sepesy Maučec, 2013).

The project dedicated almost a year to extensive evaluations involving trained subtitlers from the professional companies in the consortium, alongside automated metrics.[18] By the end of the project, subtitlers had post-edited approximately 65,000 subtitles over two evaluation phases. In the first phase, various evaluation methods were used in addition to post-editing, including rating individual subtitles on a 1–5 scale for quality, marking

recurrent errors according to a supplied taxonomy, and providing qualitative information about the subtitlers' perceptions of the process in the form of answers to open-ended questions. The second evaluation phase of the project looked at subtitlers' productivity gain from using MT, measured in terms of post-editing time. Two post-editing scenarios were examined: one with full MT output provided and one with filtered output on the basis of the application of automated quality estimation. Questionnaires were used in this evaluation phase as well, asking subtitlers to comment on their post-editing experience in terms of effort and efficiency. In both evaluation phases, subtitlers were asked to post-edit subtitles to their usual quality standards, so the resulting files would be of the same quality as subtitle files produced by professional subtitlers without the use of MT.

Subtitlers were asked to rate a total of 27,565 MT subtitles in the first phase and to post-edit them as needed to reach professional quality standards. The rating was carried out using a 1–5 scale corresponding to the amount of post-editing needed for a given subtitle, in which 1 denotes the lowest quality, i.e. an MT subtitle that is incomprehensible and requires new translation from scratch, and 5 the highest quality, with little to no editing required. Intermediate ratings ranged from 2, which indicated that significant editing effort was needed to reach publishable level, to 4 for subtitles that were generally clear and intelligible and only required minor editing; subtitles that contained various errors and mistranslations were rated 3.

The main results of this evaluation of the quality of MT output can be seen in Figure 1, which shows the percentage of subtitles assigned to each rating category overall for all language pairs.[19]

Overall, subtitlers considered 56.79% of the subtitles they rated to be of quality 4 or 5, meaning that they required little to no post-editing to be of a publishable standard. Some selected examples of the MT output which received evaluations of a 5 standard are included in Table 1.

Naturally, the biggest factor determining the success of integrating MT into the subtitling industry is whether MT delivers an improvement in productivity and efficiency.

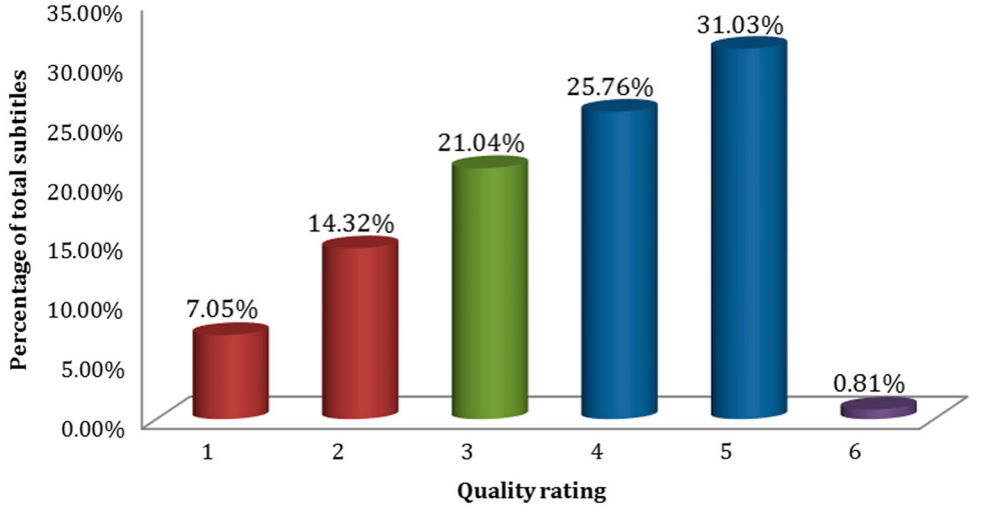

Figure 1. Global quality ratings averages.

Table 1. Machine translation examples generated by the *SUMAT* engines that scored 5 for quality.

LANG	SOURCE TEXT	LANG	MACHINE TRANSLATION	BACK TRANSLATION
EN	How long are you gonna give her a free pass?	ES	¿Cuánto tiempo le vas a dar vía libre?	How long are you gonna give her a free pass?
EN	You still don't understand, do you?	DE	Du verstehst es immer noch nicht, oder?	You still don't understand, do you?
EN	• Can you do it? • I'll have a go.	FR	• Tu peux le faire? • Je vais essayer.	• Can you do it? • I can try.
EN	Pammy, we've gotta head over to Aunt Bonnie's.	PT	Pammy, temos de ir para casa da tia Bonnie.	Pammy, we have to go to Aunt Bonnie's house.
EN	I would not ever steal from kids.	SV	Jag skulle aldrig stjäla från barn.	I would not ever steal from kids.

In the second phase of the project evaluation process, centred on productivity evaluation experiments, 19 subtitlers participated and a total of 37,104 subtitles were post-edited. The experiments focused on language pairs that would be more interesting commercially and thus language pairs that are more frequent in business terms were selected for evaluation (all out of English language pairs, and French, German, and Spanish into English), as well as the Serbian–Slovenian language pair so as to ensure the most under-resourced languages in the project received due attention. Two evaluators were used per language pair, working on six subtitle files each in total.[20] The subtitlers were asked to perform three distinct tasks: (a) translate two files directly from the template, to be used for benchmarking purposes; (b) post-edit two files in which MT output was provided for all subtitles; and (c) post-edit two filtered files in which quality estimation[21] had been applied to edit out MT subtitles below a certain quality threshold, and accordingly perform a combination of post-editing, for MT subtitles above the minimal quality threshold, and translation from scratch, for subtitles in which MT output fell below the quality threshold and was thus filtered from the files. Subtitlers were asked to record the time it took them to complete each task using a freely available and easy-to-use time-tracking tool called Toggl.[22]

The global productivity evaluation results from all the analysed language pairs in the second evaluation phase of the project are shown in Figure 2. Productivity gain/loss is expressed as the percentage of the subtitler's speed increase/decrease when post-editing the files in tasks two and three compared with his/her speed translating the corresponding benchmark file directly from the template.[23]

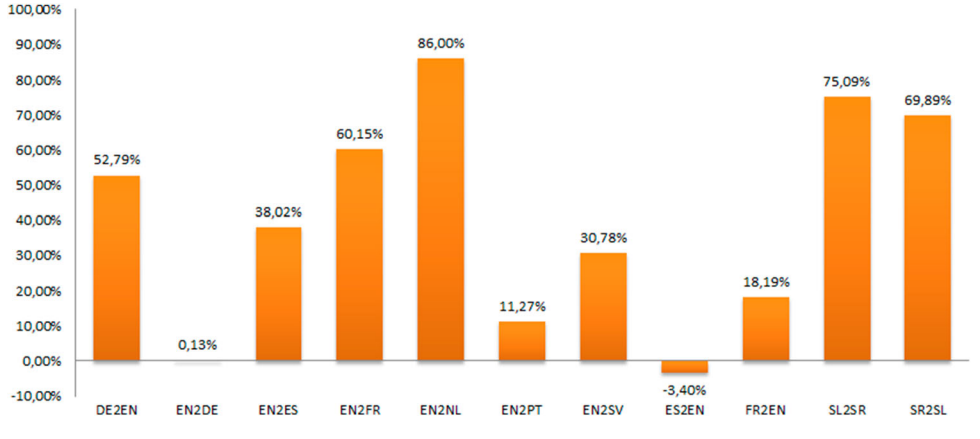

Figure 2. Average productivity gain results per language pair.

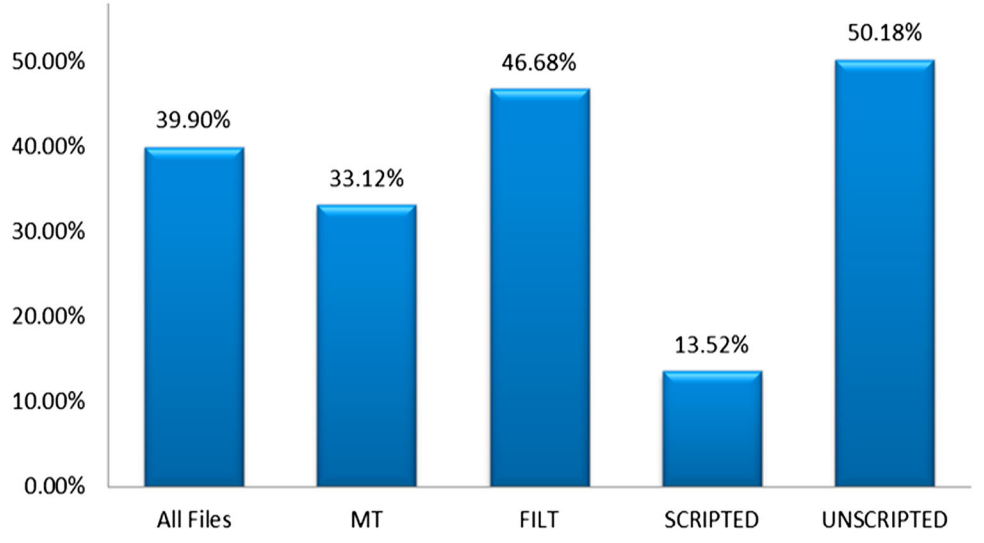

Figure 3. Global productivity gain results.

The average productivity gain achieved across all the language pairs analysed in our experiments, including filtered and unfiltered files, was 39.90%. This is a significantly higher result than our initial aim of 25% when embarking on the *SUMAT* project. Furthermore, average productivity gain in unfiltered files alone was 33.12%, compared with a 46.68% productivity gain achieved for filtered files, another significant result representing an increase of 40.92% from unfiltered to filtered files. Figure 3 summarises the global productivity results achieved in the second phase of the evaluation.

5. Machine translation for subtitling

Various factors have been shown to influence productivity gain/loss when MT is used in the translation workflow, including type of ST input, language pair, MT quality, and translator proficiency in post-editing. It is thus worth examining these factors and their interrelations in more detail in order to gain a better understanding of what MT could bring to the subtitling industry.

5.1. Strengths and limitations

A major influence on the quality of MT output and, by extension, the amount of post-editing effort and eventual productivity gain, is the type of ST input. Even with statistically trained MT engines, which can handle naturally occurring language with relative efficiency, MT faces challenges with ST that exhibit high levels of grammatical irregularity of the type often found in audiovisual texts, especially programmes that are recorded live or without a script.

Another factor that affects MT output quality is the language pair. Statistical MT is usually more successful between closely related languages such as Spanish and Portuguese, in which morpho-syntactic commonalities enable straightforward construction of

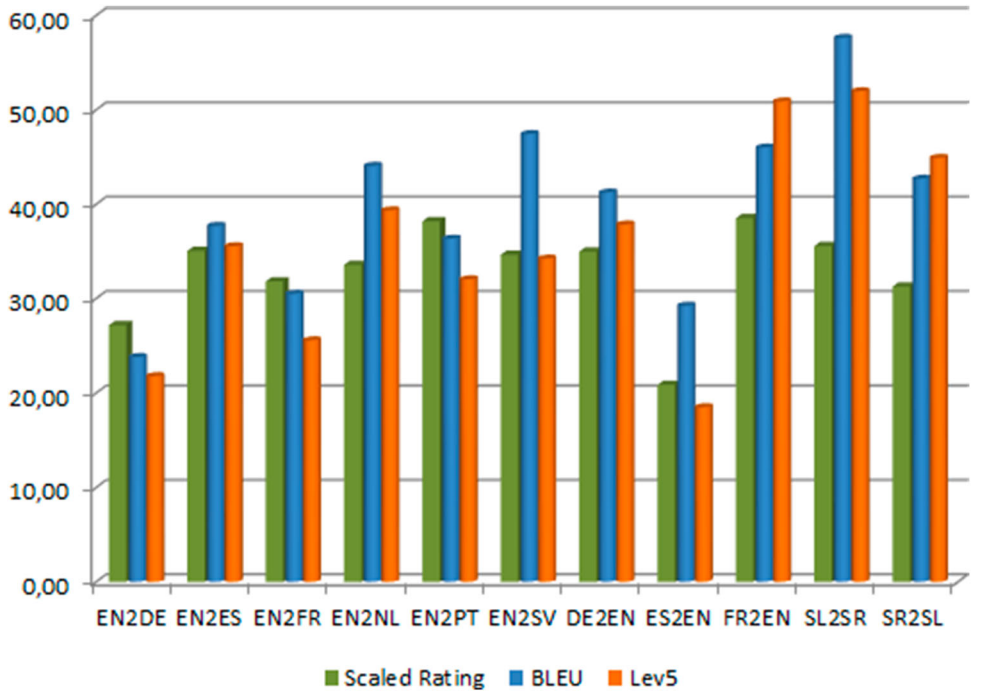

Figure 4. Human and automated evaluation metrics by language pair.

translation equivalences. Conversely, MT output quality tends to decrease for grammatically unrelated languages, which have differences in terms of morphological inflection or word order, amongst other things. Figure 4 illustrates the variability in results for the translation pairs that were part of the first *SUMAT* evaluation.

The results shown are drawn from post-edited files and based on three metrics: the average (scaled) rating assigned by the subtitlers; BLEU, one of the standard metrics in MT evaluation (Papineni, Roukos, Ward, & Zhu, 2002); and Lev5, which measures the number of MT subtitles that require five or less character-level edits to reach a reference translation (Volk, 2008, p. 10). The effect of language pair on MT output can be seen here. English to German is a notoriously difficult language direction for MT, notably due to differences in case marking, word order, and compounding. The evaluation results confirmed the difficulties for this particular language pair, which consistently scored lower than the other language pairs on the *SUMAT* test sets. The case of Spanish to English was rather surprising, as this translation pair had the highest scores on the system evaluation sets in terms of automated metrics. However, this negative result in the second phase of the evaluation shows the impact of the input files used for the assessment, as they consisted of unscripted material with large volumes of challenging ST, a factor that possibly led to this unusual result. The results were surprisingly good for the Serbian–Slovenian pair considering the low volume of data used to train the MT systems; this illustrates the impact of translating between closely related languages and indicates the quality levels that can be reached by statistical MT for under-resourced language pairs.

It is worth noting that automated quality metrics may not be the optimum tools for measuring MT output quality in the context of post-editing. Such metrics, loosely speaking, measure the quantity of differences between MT output and human reference translations, and do not indicate the variable effort needed to correct specific errors or error types. In the open subtitling domain, for instance, a mistranslated named entity (e.g. a character's full name, such as Sherry Baker, for example, being translated literally) can be more tedious to correct than a word order or agreement error, as the post-editor might need to check the ST in the former case to simply make sense of an output translation rendered incomprehensible by the translation of a proper name. Similarly, translation errors that may be perceived as easy to fix can be very labour-intensive for post-editors, even though these errors are often a complex by-product of the automated translation mechanism. A typical example of the latter would be the disappearance of source content words in the output translation.

5.2. Post-editing

As previously discussed, MT is not part of current practice in subtitling and its successful integration in the subtitling workflow depends on a number of factors related to, amongst other things, post-editing MT output.

Some of the factors that have been shown to have an effect on the successful integration of MT into the translation workflow can be grouped loosely under the umbrella of logistics. In common with other experiments, an initial productivity evaluation within the *SUMAT* pilot study (Bywood, Georgakopoulou, Volk, & Fishel, 2012) concluded that careful and extensive preparation for the post-editing task, along with thorough briefing and the opportunity to ask plentiful questions, positively influenced the post-editors' perception of their task and seemed to correlate with higher speed when post-editing. Productivity gain might thus be influenced by psychological factors, including acknowledging the importance of the post-editors' feedback and communicating the current limitations of machine translation technology. Among the feedback gathered during the evaluation campaigns of the *SUMAT* project were repeated assertions that the post-editing task became progressively easier as the subtitlers gained practice with and greater experience of typical MT errors. This feedback also provided suggestions on how to configure a user interface adapted to the optimal correction of typical simple errors such as erroneous word order or missing capitalisation, for example.

Translation experience can be viewed as a second impactful factor, with research suggesting that perhaps less-experienced and therefore slower translators may see greater increases in productivity than their more experienced colleagues (Federico, Cattelan, & Trombetti, 2012). There may be value in looking at the profiles of each of the subtitlers working on the evaluation in order to identify trends that are linked to factors such as experience, speed, and ways of working. The initial experiences in the *SUMAT* project seem to support the finding that particularly fast subtitlers are more likely to dislike and be suspicious of post-editing as they expect it to slow them down. Federico, Cattelan, and Trombetti (2012) also postulated that the individual user's user interface might have an effect on productivity gain and this factor deserves scrutiny.

Questionnaires from the evaluation in the *SUMAT* project suggested that the cognitive load of dealing with MT of poor quality was a significant factor in determining the

evaluation of the post-editing experience for the post-editor. As previously mentioned, in the final phase of the project evaluation, the *SUMAT* post-editors' productivity was also measured against files from which poor MT output was automatically removed. Although free-form feedback was mixed regarding the experience of post-editing filtered files, in this particular experiment productivity was significantly higher when filtering out 'bad' MT. It would be particularly interesting to test these findings in further studies focusing on the impact of filtering MT output to varying degrees.

Finally, a better understanding of how MT systems function and their limitations usually led to a smoother process overall. Early on in the evaluation, post-editors expressed frustration with the errors encountered, based on their (sometimes unrealistic) expectations of the performance of the MT engine and their own knowledge of the ease with which errors can be fixed by a human reviser. By educating post-editors in how MT works and also its limitations, such comments were eliminated and replaced by facts and truly useful feedback, referencing repetitive errors or specific suggestions for MT engine retraining. This education and understanding also helps lead to the realisation that MT is not a threat, but a useful tool, as shown in the post-editors' feedback: 'With shorter and simpler sentences like the ones in this episode, I think having the translation there saves quite some time' and 'There were many fixed phrases that were correct and usable'. Vastly improved tools can be made available to subtitlers, offering possible translations and concordances based on previously translated texts, and these will eliminate repetitive and time-consuming actions. This workflow is in use in the traditional text localisation industry today and, where data exist, translators show themselves to be happy with the use of MT in their day-to-day work (Guerberof Arenas, 2013).

6. Next steps

It is widely accepted that the output from MT engines is not usable for broadcast or home entertainment purposes without human intervention (Doherty, 2016, p. 958). As the TAUS (2013c, p. 47) report explains, computers 'cannot access a knowledge base that helps them decide correctly how to disambiguate a given expression in a plausible way in a given context', something which is of paramount importance for the production of high-quality subtitles. For most media, post-editing of the MT output will be necessary, therefore if we assume that the use of MT will become commonplace in the subtitling industry, then a new role is created: subtitle post-editor.

It is not customary for the translation professionals who work in the subtitling industry to have experience of translation fields in which Computer-Assisted Translation (CAT) tools and MT are commonplace. In order to address this need, it seems that existing subtitlers will need to be trained as post-editors. It has been suggested, however, that the subtitling industry could take experienced post-editors from the traditional text translation industry and train them in subtitling skills (Georgakopoulou & Bywood, 2014, p. 28); another option would be to seek out people whose skills and experience include high levels of attention to detail and possibly a tolerance for repetitive work and train them specifically in subtitle post-editing. Some research has been done using monolingual post-editors (Koehn, 2010a; Koponen & Salmi, 2015, amongst others) which shows that, somewhat counter-intuitively, post-editors with no knowledge of the source language can, in a significant number of cases, edit MT output to acceptable levels. This particular

research was carried out in the text translation domain and more work is certainly needed to quantify the necessary skillset for good subtitle post-editors.

If MT is to be successfully adopted, trained post-editors will be necessary in the subtitling industry, and this fact then highlights a significant skills gap due to the comparatively late adoption of this technology. Pym (2012, p. 15) points to the similarities between post-editing and revision work and perhaps these may prove to be a helpful resource when teaching post-editing skills. Revision in the subtitling industry is typically undertaken by more experienced subtitlers, whose job is to perform a thorough check of a colleague's work for translation accuracy as well as grammar and style, and some of these skills are undoubtedly transferable. What has become apparent from our experiments and others like them, however, is that while translation and revision require some similar skills, the problems that occur in a human-translated text are not comparable to those which arise in MT output. For MT to work in subtitling, the necessary post-editing skills need to be acquired or taught. Although specific errors produced by statistical translation engines will not always be fully predictable, post-editing practice and training provides an increased ability to recognise and expect typical errors produced by MT systems. Indeed, some initial evidence from the *SUMAT* post-editors indicates that they did observe an increase in the speed of their work during the evaluation, probably as a result of having had experience of the types of mistakes to expect of a machine, which then facilitated decisions on what and how to post-edit.[24] This should be further investigated, as other research has shown that, at least from a subjective point of view, this is not the case (Guerberof Arenas, 2013).

Postgraduate AVT courses are slowly beginning to see the need for post-editing, however provision is still scarce and resources need to be devoted to this area of translator training. In an attempt to counter this lack of higher education provision, individual language service providers are drafting their own post-editing guidelines[25] and industry bodies such as TAUS (2013a, 2013b) are addressing this issue. An encouraging sign is the emergence of modules and seminars in higher education institutions offering an overview of post-editing[26] and partnerships between MT developers and translator training courses.[27]

In addition to the necessity for training at higher education level, an open dialogue between subtitle translators working as post-editors and developers of MT technology is necessary in order to adapt technology to the needs of its users. This will encourage the translation of scientific knowledge to daily practice and inform further research in this area. Feedback from post-editors, which will vary across language pairs, serves as valuable data for the continuous improvement of the MT systems and their user interfaces. Such input, which can comprise linguistic analysis and refinement of training data, as well as editing of MT errors and identification of patterns, is crucial in eliminating or minimising mistakes in the MT output. As Pym (2012, p. 15) states, the translators themselves are the ones best suited to 'investigate the human aspects of translation technology and hence the ones that pinpoint more easily what it is that can make this technology truly revolutionary'.

7. Conclusions

In conclusion, the growth in content to be subtitled, coupled with the drastic fall in available revenue, has meant that the subtitling industry faces a situation in which it can no longer

function using the production methods it currently employs. The localisation industry at large has embraced translation automation and, consequently, the subtitling industry has the opportunity to learn from this experience. The results of the *SUMAT* project corroborate that MT provides a promising option for partially automating the subtitling workflow and could result in the productivity gains the industry requires. Various factors affecting productivity when using MT have been discussed above, from the need to filter poor MT output to the perception of the post-editing task by professional subtitlers; further studies will be necessary for a detailed assessment of each factor's scope and weight. The *SUMAT* project has built SMT systems tuned on subtitles and conducted a large-scale evaluation using both human and automated metrics. The results of this evaluation proved promising in terms of MT quality and productivity gains, both crucial factors in determining whether this technology will be successfully adopted in the subtitling industry.

It is important to stress that MT systems for use in subtitling are not intended to replace humans, but rather as an aid to productivity to address some of the challenges outlined in this paper. With this in mind, it is apparent that a new job profile for the industry is emerging, that of subtitle post-editor; education and training for this role will require close collaboration between the industry, professional subtitlers, MT researchers, and higher education providers.

Notes

1. http://www.ethanzuckerman.com/blog/the-polyglot-internet
2. http://youtubeukblog.blogspot.co.uk/2008/11/auto-translate-now-available-for-videos.html
3. SUMAT (An Online Service for Subtitling by Machine Translation). Retrieved from www.fp7-sumat-project.eu
4. Interlingual subtitling: subtitling from one language to another, as opposed to intralingual subtitling, or subtitling in the same language as the film audio, which mainly enables accessibility for the deaf and hard of hearing audience.
5. www.translatetv.com
6. MUSA (Multilingual Subtitling of multimedia content). Retrieved from http://sifnos.ilsp.gr/musa/
7. eTITLE (2004). Universitat Pompeu Fabra, Barcelona. Retrieved from https://portal.upf.edu/web/glicom/e-title
8. http://www.deluxemediaeurope.com
9. http://www.vsi.tv/
10. http://www.subtitling.net/
11. http://www.ericsson.com/broadcastandmedia/locations/germany-berlin/
12. www.capitatranslationinterpreting.com
13. www.textshuttle.ch
14. www.atc.gr
15. www.vicomtech.es
16. www.um.si
17. The two main additional corpora were OpenSubtitles, which includes large amounts of crowd-sourced subtitles, and the Europarl corpus (Koehn, 2005), extracted from the proceedings of the European parliament; all additional corpora were collected from the OPUS repository (Tiedemann, 2012). See the project's final report for more details: http://www.fp7-sumat-project.eu/final-sumat-report-and-findings-available-to-download/index.html
18. For a detailed description of these evaluations and their results, see Etchegoyhen et al. (2014).
19. Category 6 refers to rating errors, mostly unrated MT subtitles or typographical errors in the rating.

20. The Serbian–Slovenian language pairs were an exception, in which, due to logistical issues, only one evaluator per translation direction was used, working on three files in total per pair.
21. Quality estimation was performed using the QuEst toolkit (Specia, Shah, de Souza, Cohn, & Kessler, 2013). We used the quality annotation data collected in the first phase of the evaluation to train and test quality estimation models using QuEst baseline features, which cover source and target sentence properties such as length, language model probabilities, and average number of translations per word, among others.
22. http://www.toggl.com. Although the tasks were time-measured, subtitlers were instructed to work according to their usual practice; that is, to take breaks if that was normal for them so as to ensure the evaluation conditions would be comparable to their actual working conditions. In such cases, they were asked to record the time each task took them, excluding the breaks they took.
23. The results presented here slightly differ from the ones in Etchegoyhen et al. (2014), with a change from 38.23% to 39.90% in terms of global productivity, for instance. This is due to one EN2NL file, the final results of which were included at a later stage, resulting in slightly greater productivity results.
24. Cf. Wendt (2008) for the impact of post-editor training on productivity gains in the translation of Microsoft's Knowledge Base articles.
25. Cf. Beregovaya, Lavie, Clarke, and Denkowski (2013).
26. See, for example, the module on Machine Translation and Post-Editing offered by the University of Bologna (http://www.unibo.it/en/teaching/course-unit-catalogue/course-unit/2015/397338).
27. https://www.gala-global.org/publications/dcu-and-ucl-post-graduates-get-hands-latest-machine-translation-technology

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the European Commission (EU) under ICT Policy Support programme 2011–2014, grant agreement 270919.Seventh Framework Programme

Notes on contributors

Lindsay Bywood is Senior Lecturer in Translation Studies at the University of Westminster and teaches translation, audiovisual translation, and project management for translators at postgraduate level. She holds a PhD in subtitling from UCL, an MA in German and Philosophy from the University of Oxford, and an MA in Translation from the University of Salford. Her research centres around the diachronic variation in the subtitling of German films into English, with other research interests in machine translation and post-editing, accessibility, and the interface between industry and academia. Before becoming an academic she worked for many years in the audiovisual translation industry and is now responsible for the postgraduate professional development programme for translators and interpreters.

Panayota (Yota) Georgakopoulou holds a PhD in translation and subtitling and is a seasoned operations executive, currently serving as Senior Director, Research & International Development, at Deluxe Media. She leads translation initiatives as well as research on language technologies and tools, and their application in subtitling. Starting off as a translator and subtitler, Yota taught translation and set up subtitling modules at two UK universities, and went on to manage a small subtitling company for several years. She has also served as external adviser and external examiner

for postgraduate programmes in audiovisual translation, and in the steering and advisory committees of international conferences.

Thierry Etchegoyhen is a Principal Researcher in the Human Speech and Language Technology Group (HSLT) at Vicomtech-IK4. He holds a PhD in Computational Linguistics from the University of Geneva and has worked over the years as both an academic researcher and a developer of natural language processing products. His activities have covered research and development for core language processing components as well as the transfer of advanced technology to the industry. He currently leads machine translation research and development in the HSLT group and has been active in several national and European projects in the field.

ORCID

Lindsay Bywood http://orcid.org/0000-0002-1860-6647

References

Armstrong, S., Caffrey, C., & Flanagan, M. (2006). Translating DVD subtitles from English-German and English-Japanese using example-based machine translation. In *MuTra 2006 –Audiovisual Translation Scenarios: Conference Proceedings*. Copenhagen. Retrieved from http://www.translationconcepts.org/pdf/MuTra_2006_Proceedings.pdf

Barbour, M. (2013). Dell & Welocalise: the biggest machine translation programme ever? Retrieved from https://www.welocalize.com/dell-welocalize-the-biggest-machine-translation-program-ever/

Bender, O. (2010). *Robust machine translation for multi-domain tasks* (Unpublished PhD dissertation). Aachen: RWTH Aachen University.

Beregovaya, O. Lavie, A. Clarke, D., & Denkowski, M. (2013). Analyzing and predicting MT utility and post-editing productivity in enterprise-scale translation projects. In *Proceedings of Machine Translation Summit XIV*. Switzerland: EAMT.

Bywood, L. Georgakopoulou, P. Volk, M., & Fishel, M. (2012). *What is the productivity gain in machine translation of subtitles?* Paper presented at Languages & The Media, Berlin.

Castilho Monteiro de Sousa, S., Aziz, W., & Specia, L. (2011). *Assessing the post-editing effort for automatic and semi-automatic translations of DVD subtitles*. Paper presented at Recent Advances in Natural Language Processing Conference, Hissar, Bulgaria.

Doherty, S. (2016). The impact of translation technologies on the process and product of translation. *International Journal of Communication*, 10, 947–969. Retrieved from http://ijoc.org/index.php/ijoc/article/viewFile/3499/1573

Etchegoyhen, T., Bywood, L., Fishel, M., Georgakopoulou, P., Jiang, J., van Loenhout, G., … Volk, M. (2014). Machine translation for subtitling: a large-scale evaluation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, 46–53. Reykjavik. Retrieved from http://www.lrec-conf.org/proceedings/lrec2014/pdf/463_Paper.pdf

Etchegoyhen, T., Fishel, M., Jiang, J., & Sepesy Maučec, M. (2013). SMT approaches for commercial translation of subtitles. In *Proceedings of Machine Translation Summit XIV*, Nice, 369–370. Retrieved from http://www.mt-archive.info/10/MTS-2013-Etchegoyhen.pdf

Federico, M., Cattelan, A., & Trombetti, M. (2012). Measuring user productivity in machine translation enhanced computer assisted translation. In *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA)*. Madison, WI: AMTA.

Georgakopoulou, P. (2006, July 6). Subtitling and globalisation. *The Journal of Specialised Translation*. Retrieved from http://www.jostrans.org/issue06/art_georgakopoulou.php

Georgakopoulou, P., & Bywood, L. (2014, January/February, 24–28). Machine translation in subtitling and the rising profile of the post-editor. *MultiLingual*.

Graham, P. (2005). Web 2.0. Retrieved from http://www.paulgraham.com/web20.html

Guerberof Arenas, A. (2013, January 19). What do professional translators think about post-editing? *The Journal of Specialised Translation*. Retrieved from http://www.jostrans.org/issue19/art_guerberof.pdf

Hardmeier, C., & Volk, M. (2009). Using linguistic annotations in statistical machine translation of film subtitles. In *Proceedings of 17th Nordic Conference on Computational Linguistics (Nodalida)*, 57–64. Linköping, Sweden: Linköping University Electronic Press.

House of Lords Select Committee on Communications. (2011). Regulation of television advertising: 1st report of session 2010-11, report with evidence, *Volume 99 of HL (Series) (Great Britain. Parliament. House of Lords)*. London: The Stationery Office.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of Machine Translation Summit X. Phuket*, 79–86. Thailand.

Koehn, P. (2010a). Enabling monolingual translators: Post-editing vs. options. In *Proceedings from the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*, 537–545. Los Angeles, CA: Association for Computational Linguistics.

Koehn, P. (2010b). *Statistical machine translation*. Cambridge: Cambridge University Press.

Koponen, M., & Salmi, L. (2015, January 23). On the correctness of machine translation: A machine translation post-editing task. *The Journal of Specialised Translation*. Retrieved from http://www.jostrans.org/issue23/art_koponen.pdf

Matusov, E., Ney, H., Popovic, M., & Zens, R. (2004). Statistical machine translation of spontaneous speech with scarce resources. In *Proceedings of the International Workshop on Spoken Language Translation*, 139–146.

McCarthy, M., & Carter, R. (2001). Ten criteria for a spoken grammar. In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 51–75). Mahwah, NJ: Lawrence Erlbaum Associates.

Melero, M., Oliver, A., & Badia, T. (2006). Automatic multilingual subtitling in the eTITLE project. In *Proceedings of ASLIB Translating and the Computer 28*. London: ASLIB.

O'Hagan, M. (2003). Can language technology respond to the subtitler's dilemma? - A preliminary study. In *Proceedings of ASLIB Translating and the Computer 25*. London: ASLIB.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318.

Paterson, K. Caygill, C., & Sewell, R. (2012). *A handbook of spoken grammar. Strategies for speaking natural English*. London: Delta Publishing.

Piperidis, S., Demiros, I., & Prokopidis, P. (2005, January 24-28). Infrastructure for a multilingual subtitle generation system. In *9th International Symposium on Social Communication*, Santiago de Cuba, Cuba. Retrieved from http://sifnos.ilsp.gr/musa/publications/Infrastructure%20for%20a%20multilingual%20subtitle%20generation%20system-final.pdf

Piperidis, S., Demiros, I., Prokopidis, P., Vanroose, P., Hoethker, A., Daelemans, W., … Karavidas, Y. (2004). Multimodal multilingual resources in the subtitling process. In *Proceedings of the 4th International Language Resources and Evaluation Conference*. Lisbon: ELRA.

Popowich, F., Mcfetridge, P., Turcato, D., & Toole, J. (2000). Machine translation of closed captions. *Machine Translation*, 15, 311–341. Retrieved from http://www.jstor.org/stable/20060451?seq=1#page_scan_tab_contents

PROMT. (2012). *PROMT Translates TripAdvisor's Reviews*. Retrieved from http://www.promt.com/media/news/44383/

Pym, A. (2012). Democratizing translation technologies—the role of humanistic research. In V. Cannavina & A. Fellet (Eds.), *Language and translation automation conference* (pp. 14–29). Rome: The Big Wave.

Reid, A. (2013). *Introduction to machine translation and the post-editing paradigm shift*. Ebook, SDL. Retrieved from file:///H:/Publications/SDL_eB_Shaping_The_Future_Of_Content_Translation_hires_tcm73-58317.pdf

Specia, L., Shah, K., de Souza, J. G., Cohn, T., & Kessler, F.B. (2013). QuEst–a translation quality estimation framework. In *Proceedings of the 51st ACL: System Demonstrations* (pp. 79–84). New York: Curran Associates.

SUMAT (An Online Service for SUbtitling by MAchine Translation). Retrieved from www.fp7-sumat-project.eu

TAUS. (2013a). *Best practice guidelines. Quality evaluation using adequacy and/or fluency approaches*. Retrieved from https://www.taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluency-guidelines

TAUS. (2013b). *Best practice guidelines. Quality evaluation using an error typology approach*. Retrieved from https://www.taus.net/academy/best-practices/evaluate-best-practices/error-typology-guidelines

TAUS. (2013c). *Translation technology landscape report*. Retrieved from https://www.taus.net/think-tank/reports/translate-reports/taus-translation-technology-landscape-report

Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC2012)*. Istanbul, Turkey.

University of Edinburgh. (2014). *The moses machine translation toolkit, REF2014 impact case study*. Retrieved from http://impact.ref.ac.uk/casestudies2/refservice.svc/GetCaseStudyPDF/23947

Volk, M. (2008). The automatic translation of film subtitles. A machine translation success story? In J. Nivre, M. Dahllöf, & B. Megyesi (Eds.), *Resourceful language technology: Festschrift in honor of Anna Sågvall Hein* (pp. 202–214). Uppsala: Acta Universitatis Upsaliensis.

Wendt, C. (2008, October 21–25). *Large-scale deployment of statistical machine translation: Example Microsoft*. Paper presented at AMTA, Waikiki, HI.

Wohlsen, M. (2014, April 29). The next big thing you missed: why ebay, not google could save automated translation. *Wired*. Retrieved from http://www.wired.com/2014/04/the-next-big-thing-you-missed-how-ebay-can-change-online-translation/#slide-1

YouTube Official Blog. (2008). *Auto-translate now available for videos with captions*. Retrieved from http://youtubeukblog.blogspot.co.uk/2008/11/auto-translate-now-available-for-videos.html

Zuckerman, E. (2008). *The polyglot Internet*. Retrieved from http://www.ethanzuckerman.com/blog/the-polyglot-internet