ATLANTIS
Journal of the Spanish Association of Anglo-American Studies
44.1 (June 2022): 233-244
e-ISSN 1989-6840
DOI: http://doi.org/10.28914/Atlantis-2022-44.1.13

Machine Translation and the Assessment of Translation Quality. A Critical Review of

Bernard Scott, ed. 2018. *Translation*, *Brains and the Computer*. Cham: Springer. xvi + 241 pp. ISBN: 978-3-030-09538-3.

Joss Morkens, Sheila Castilho, Federico Gaspari and Stephen Doherty, eds. 2018. *Translation Quality Assessment*. *From Principles to Practice*. Cham: Springer. xii + 287 pp. ISBN: 978-3-030-08206-2.

Elisabet Comelles Pujadas
Universitat de Barcelona
elicomelles@ub.edu

In the globalised world we are living in, the need to quickly access and exchange information in different languages is vital. The enormous amount of information that is shared and exchanged by millions of people around the world needs to be made available in different languages to guarantee democratic access to it. Thus, the fast and trustful translation of that information is essential. That is the reason why Machine Translation (MT) has played a key role in recent decades, especially since companies like Google made their MT engines available globally.

The books reviewed in this article offer the reader a complete picture of the field of MT and its evaluation. The first book, *Translation*, *Brains and the Computer*, edited by Bernard Scott (2018), approaches MT and the traditional problems linked to the field of Natural Language Processing (NLP) from a psycholinguistic and neurolinguistic perspective. It also describes the Logos Model (LM) MT system, which uses English as its source language, and compares it to other MT technologies. The second book, *Translation Quality Assessment*. *From Principles to Practice*, edited by Joss Morkens et al. (2018), complements the first in dealing with the assessment of both Human and MT Quality, trying to bring principles and practice together in the academic as well as institutional and industry settings.

MT involves the automatic translation of a text from a source language into a target language. This is one of the most complex and challenging tasks in the field of NLP because it involves most of the types of knowledge that humans possess (i.e., grammar, semantics, knowledge of the world, etc.). According to Basnett, when a person translates a text "a process of decoding and encoding takes place" (1980, 24); that is, the translation process involves decoding the meaning of the source text and encoding this meaning into the target language, which is a complex cognitive operation. Decoding a source text therefore means that the translator must understand and analyse the source text in its entirety, which requires good knowledge of all dimensions of the source language (e.g., lexicon, grammar, semantics) as well as knowledge of the source culture. In addition, the process of encoding also implies having the same knowledge of the target language. Reproducing such a complex operation is the challenge of MT. As pointed out by Hutchins and Sommers (1992, 2), "[the] major obstacles to translating by computer are, as they have always been, not computational but linguistic. They are the problems of lexical ambiguity, of syntactic complexity, of vocabulary differences between languages, of elliptical and 'ungrammatical' constructions, of, in brief, extracting the 'meaning' of sentences and texts from analysis of written signs and producing sentences and texts in another set of linguistic symbols with an equivalent meaning."

The idea of using a computer to translate natural languages dates back to 1946 when Warren Weaver wrote a memorandum where he addressed the future prospects of MT and suggested several methods to approach it. Later on, in the 1950s and 1960s the number of universities, research groups and companies interested in MT rocketed, particularly in the USA and the Soviet Union. It was indeed in the late 1960s when Bernard Scott and his group started working on their MT system, the Logos Model (LM). In the history of MT, various approaches have been deployed, the most well-known being Rule-Based MT systems (RBMT), where transfer rules are applied to transform an input sentence into a target one; Statistical MT (SMT) systems (Kohen 2009), which build probabilistic models of faithfulness (meaning) and fluency and use these models to choose the most probable translation; and Neural Machine Translation (NMT) systems (Kohen 2020), which are based on neural networks that try to simulate human neurons. Both SMT and NMT systems need large amounts of data for training, whereas RBMT needs a trained linguist to write the rules. The MT system presented in Scott's (2018) book is different from the above though, as explained in the second chapter of his book where he describes the circumstances that brought LM into existence, it introduces the psycholinguistic and neurolinguistic principles in which the model is inspired and gives an overview of the system translation process. Indeed, a comparison between LM and SMT is also provided which emphasises the fact that, in contrast to RBMT, both SMT and LM systems are pattern-driven. In other words, although SMT is grounded in statistical data and LM in linguistic understanding, both translation processes are based on patterns of language rather than on a set of rules. The last section of this chapter looks into the similarities between NMT and Logos. The author emphasises that, despite the obvious differences (e.g., LM is linguistically-based), the architecture and basic functionality of the two systems are comparable. This topic is further addressed in chapter eight.

As already mentioned at the beginning of this review, the originality of this book lies in its combined psycholinguistic and neurolinguistic approach. Scott (2018) firmly believes that understanding how the brain processes, learns and acquires language, and applying that knowledge to MT will definitely help achieve high-quality MT output and solve two common problems: a) ambiguity and b) complexity. These two linguistic issues, already identified by Hutchins and Sommers (1992), are tackled in chapters three and four of the book. Scott (2018, 41) defines *ambiguity* as "a linguistic situation capable of more than one interpretation" and highlights the fact that, unlike humans, MT systems are not capable of dealing with ambiguity, and uses this to reinforce the idea that the processes of language understanding and translation should be included in MT systems. He provides examples of how Logos, which has been developed following those principles, is in fact able to cope with ambiguity when translating from English to German or French. As for the complexity issue, Scott (2018) coins the term *Cognitive Complexity* to refer to the difficulty developers experience in maintaining complex systems, which he claims particularly affects RBMT, whereby adding a new rule to resolve an issue might undo the resolution of another issue. However, I would like to stress that SMT and NMT do not escape this same problem. NMT systems are a black box, so knowing why the system has made certain word choices is a mystery (Kohen and Knowles 2017). Scott (2018) believes that the ability of a system to deal with complexity issues is directly related to its potential for high-quality translations, thus he seeks to explore why the brain seems to be free of complexity issues when handling language. In this line, chapter five examines several approaches that try to account for the brain mechanisms that decode individual words into unambiguous meaningful sentences, focusing particularly on how syntax and semantics relate to each other. The views thoroughly examined by the author are Generativism and Cognitive Linguistics (i.e., constructionists). The constructionists' approach definitely matches the author's view on MT and the way the semantico-syntactic grammar of the Logos Model is structured, considering syntax and semantics as an inseparable continuum. This idea of syntax and semantics being interrelated is also backed by several neurolinguistic studies (Opitz and Friederici 2007; Zhang et al. 2010; Duff and Brown-Schmidt 2012; Kurczek et al. 2013; Bonhage et al. 2015), as well as an experiment carried out by the author's team. The experiment consists in translating the ungrammatical sentence *The weather bureaus all reports that the winter will be mild* into French. Several MT systems are used (namely, Google SMT Translate, Google NMT Translate, Bing SMT, Bing NMT, PROMT Translator, SYSTRANNet,[1] LISA Lab NMT and the Logos Model). Interestingly, only

[1] The PROMT and SYSTRANNet systems used in the experiment reported were a hybrid linguistic / statistical system. Nowadays the NMT technology is available.

Google NMT, Lisa Lab NMT and the Logos Model translate *reports* as a verb in French and successfully resolve the agreement with the plural subject. This experiment seems to indicate that MT systems that somehow try to emulate the workings of the brain can successfully overcome syntactic issues thanks to the prevalence of semantics; in other words, syntactic violations do not impede semantic comprehension. It should be noted, however, that in the case of NMT systems their performance strongly depends on the data they have been trained on, whereas with Logos, its excellent way of handling that ungrammatical sentence is due to its use of the SAL language,[2] a taxonomic semantico-syntactic language which integrates syntax and semantics, just as the brain seems to do. SAL, which has only been developed for English and German,[3] is organised into supersets, sets and subsets. It consists of approximately 1,000 elements for all Parts of Speech (PoS) combined. The English variant of SAL is fully described in chapter nine, where an inventory of SAL PoS, supersets, sets and subsets is detailed.

The Logos Model radically differs from traditional approaches (RBMT) in which the syntactic and semantic information are kept separate. LM consists of a lexicon where lexical entries are represented by means of SAL and a database where pattern-rules are stored. These pattern-rules have an optional target-action component that links to target generation. Thus, as is explained in chapter six, the process of translation with the Logos Model is as follows: the input string is first transformed into a SAL input string and then the SAL input string is matched to a stored SAL-pattern rule. Unlike RBMT, this matching is input-driven; in other words, the input stream looks for a relevant match in a collection of pattern-rules. Once the match is performed, the target action uses the semantic information of the source pattern to transform the source syntax into an appropriate target order ready for the final literal translation.

The next chapter moves away from Logos' internal workings and explores the limits of translation quality. It is commonly believed that in MT the structure of the target sentence is largely determined by the structure of the source sentence. Needless to say, that is unthinkable when it comes to human translators. Actually, translating from one language to another quite frequently involves structural shifts away from the source language syntax. In this chapter, Scott (2018) examines several examples to check if MT systems can imitate human translators in this regard. To this end, several English sentences showing problematic translation issues related to shifts (e.g., translation of certain infinitive clauses) are translated into German, and the MT output provided by several systems (i.e., Google SMT Translate, Systranet, Logos Model, Google NMT Translate and Bing NMT Translator) is discussed. The results show that SMT and NMT systems seem to perform the desired shifts, which might to some extent be explained by the fact that they are trained on human translations. On the other hand, RBMT systems do not handle the translation as successfully as data-driven systems, seeming to be more restricted by the syntax of the source language.

[2] Semantico-Syntactic Abstraction Language.

[3] In the experiment reported, only the English version of SAL was used.

In chapter eight, Scott (2018) puts forward some more ideas on how to improve MT by following the language processing mechanisms in the brain. He focuses on two key learning processes, generalisation and abstraction, and states that grammar rules seem to have both qualities. Notice that although the author is quite critical of Chomsky's and Generativists' views, he accepts that Chomsky's turn to syntax was of value at the outset of MT. As the author affirms, "[this] reduction of literal sentences to abstract, generalized syntactic strings made MT possible" (2018, 184). RBMT systems were the state-of-the-art technology in the 1980s and 1990s, but their limitations made developers turn to SMT and NMT systems. However, these data-driven non-linguistic systems suffer from their lack of generality; in other words, they have problems dealing with text that differs from the corpora they have been trained on. At this stage, the author seems to have revealed a paradox: MT models that do not use rules fail in linguistic competence, but those that use them show complexity issues that affect their potential. Scott's (2018) proposal to overcome that paradox is, yet again, relying on how the brain processes language, this time by looking into the concept of *complementary learning*. To do so, he refers to Kumaran et al. (2016), where the authors state that the brain must have two complementary systems of learning, one that records the immediate specifics of experience and another one that brings about more reflective learning. This complementary learning ensures that the new experience learned is accommodated with the old one, avoiding the risk of undoing something that has already been learnt. According to Scott (2018), for MT learning to be unlimited, new learning must not cause any disruptions in the current capability of the system. Therefore, MT learning also needs to be complementary.

The second book considered in this review, *Translation Quality Assessment. From Principles to Practice*, deals with the evaluation of translation quality in detail. The book is divided into three parts: part one provides an overview of the state-of-the-art in Translation Quality Assessment (TQA), part two deals with the development of applications of TQA and part three explores empirical studies of new applications of TQA.

Part one starts with Castilho et al.'s critical overview of how TQA is defined and measured in both human and MT workflows, in research, education and industry and it also addresses the issue of TQA's goal being different depending on the context. The authors walk the reader through several widely-extended TQA methodologies, discussed later in Lomel's chapter, and examine the most well-known human TQA approaches (e.g., adequacy and fluency, ranking, acceptability, etc.) and automatic ones (i.e. evaluation metrics). They also refer to some TQA issues, namely the lack of standardisation and its inherent inconsistency, and raise awareness of the outsourcing of translation in the industry—which is even happening in the European Union Directorate-General for Translation (EUDGT)—which affects both process and product quality. This issue is tackled in detail in the following chapter by Drugan et al., which discusses how Translation Quality and Translation Quality Management are dealt with in the EUDGT, where more than eighty per cent of the documents are translated from English. The authors highlight that translating texts in the EU is quite a sensitive issue because: a) all translations are legally equivalent and equally authentic; and b) the number of outsourced translations is growing. As a consequence, the procedures in the translation workflow must be very well defined and closely observed to guarantee high-quality accurate translations. The authors guide the reader through the tools, resources and processes that help maintain consistency between translations and ensure high-quality translated texts.

In the last decade, a new trend in translation and the assessment of translation quality has become very popular both in industry and research: *crowdsourcing*. Jiménez-Crespo examines this new method in the third chapter. In a previous study (Jiménez-Crespo 2017, 76) the author defines *translation crowdsourcing* as "the collaborative translation process performed through dedicated web platforms that are initiated by companies or organisations and in which participants collaborate with motivations other than strictly monetary." This new form of translation has a clear impact on the quality of the translation, and can vary depending on the content, the type of text and especially the skills of the participants. Thus, in an increasingly digitalised world where the need for translated content is constantly growing, the quality of a translation does not seem to depend so much on the text itself, as in the scenarios referred to in Scott's (2018) book, but on other variables such as the purpose of the text, the skills of the participants and whether there is a process of revision. Jimenez-Crespo also emphasises that crowdsourcing has changed the responsibility for translation quality, which has traditionally fallen on the translator and now shifts to the client. Clients decide how much they are willing to pay for the translation, so the service they receive depends on that. Crowdsourcing has been extensively used in MT research as a way to improve and train MT systems, as well as to evaluate MT output. In the last decade, several studies have explored the use of crowdsourcing to evaluate MT output instead of using automatic metrics or human evaluators (Goto et al. 2014). These studies have shown that crowdsourcing can be used at a system level (i.e., comparing different MT systems) but it does not work at the sentence level, where volunteers do not achieve the same results as professional translators.

Part one finishes with the chapter "On Education and Training in Translation Quality Assessment," which differs considerably from the rest of the book in that it explores how training in TQA is carried out in academia. The authors feel that TQA is still overlooked in the academic context and that including TQA-related content on the syllabus of translation courses will benefit both students and the industry. Although most university degrees and masters related to translation do include courses on translation technologies (e.g., translation memories, MT and Post-editing [PE]), TQA models and tools are not sufficiently covered, if, that is, they are not completely disregarded. The authors refer to a study carried out by Huertas Barros and Vine (2017), who found that, worryingly, most of the UK universities surveyed did not believe that contemporary industry TQA was relevant in the academic context. Unfortunately, this provides irrefutable evidence of the still undeniable gap between academia and industry, as well as the great need for this gap to be bridged.

The second part of this book covers some of the applications of TQA. It starts with Lomel's chapter "Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies," which takes up on the issue of the lack of standardisation in TQA already introduced in part one. The author presents the efforts that have been made to overcome subjectivity and the lack of standardisation in TQA and introduces three error typologies: the Multimodal Quality Metrics (MQM) Framework, the TAUS Dynamic Quality Framework (DQF) Error Typology and the harmonised DQF-MQM Error Typology, which blends the previous two. These frameworks respond to the pressing need for TQA standardisation in both academia and industry. Living proof of that is the considerable interest that the European-funded harmonised typology has generated in the industry.

The topic of error typology and error analysis is also discussed in Mayya Popović's chapter, although unlike Lomel, she focuses on MT output only. The author thoroughly examines several error typologies that have been developed throughout the years using datasets that had English as the source language. It is well known that manual error classification provides a detailed analysis that can be of enormous help to improve MT systems; however, it is time-consuming, expensive and often lacks consistency among annotators. In order to address these issues, the author suggests two solutions: a) a general error typology that goes from broader classes to more specific issues; and b) the use of automatic error classification and PE. Popović introduces two of the most commonly used automatic error classification metrics that compare MT output with references: Hjerson+ (Popović et al. 2015) and Addicter (Fishel et al. 2011; Zeman et al. 2011). As she points out, although these metrics are faster, cheaper and more consistent than manual error classification, they cannot cover all the issues, hence the need for further research on this topic.

The next chapter, "Quality Expectations of Machine Translation" by Andy Way, gives an overview of the current use of MT and how it is evaluated. Way coincides with Jimenez-Crespo in believing that MT quality evaluation should consider the actual use-case and usage of the translation. The author also reviews several methods used to evaluate MT output, both human-based and automatic (i.e., similarity metrics that compare MT output with reference translations). Human evaluation has been considered time-consuming, expensive and inconsistent, but MT metrics also show some other drawbacks, as discussed in Popović's chapter. In addition to the downsides put forward by Popović, Way mentions others, such as the fact that MT output is quite often compared to a single reference translation, and the bias that some MT metrics show—some seeming to prefer shorter sentences, whereas others prefer longer ones. He also raises awareness of correlation with human judgements; for example, the widely-used metric BLEU (Papineni et al. 2002) has shown poorer correlation with human judgements when compared to other linguistically-informed metrics; however, BLEU is the most popular metric in the area because it is fast and easy to use.

Way agrees with Scott (2018) that MT quality has improved in recent years thanks to new translation technologies such as NMT, which has been demonstrated to reach very high levels of quality in comparison to other types of systems (e.g., statistically-based systems). However, the substantial improvements in NMT do not seem to be captured by traditional MT metrics that operate at the n-gram level. Therefore, there seems to be an urgent need for new metrics that are capable of estimating for NMT performance.

Following on Way's and Jimenez-Crespo's view that TQA should consider the use and purpose of the text translated, the chapter by Doherty and Kruger discusses the assessment of translation quality in audiovisual translation (AVT), particularly the assessment of human and machine-generated subtitles and captions. In AVT the use case becomes extremely relevant. The audience's needs and expectations, as well as the place and medium in which the viewing occurs, are very much taken into account. AVT has been traditionally assessed on the basis of accuracy, presentation and timing, although recently metrics such as the Net Error Rate (NER) Model (Romero-Fresco and Pérez 2015) have been introduced. These metrics are similar to traditional TQA metrics, but they do not seem to account for all the typical AVT error types (Mikul 2014). Consequently, as the author points out, researchers are working intensively in order to better understand how viewers process and receive audiovisual content in an attempt to better assess translation quality in this context. This research is grounded in cognitive aspects that focus on visual attention and the Cognitive Load Theory (Plass et al. 2010).

Part three of the volume consists of three chapters that analyse how TQA is brought into practice. The first discusses the applications and future of MT Quality Estimation (QE), while the second and the third look into how MT and TQA can be applied to less common areas: respectively, using MT and PE to help non-native speakers of English write academic papers, and employing MT to translate literature.

Specia and Shah discuss the use of QE and its perspectives. Although this chapter has been included in the third part of the book, its topic seems to be more relevant to part two, since QE is another approach to TQA, and is commonly contrasted with MT metrics. The goal of QE is to estimate how good a translated text is without relying on the comparison with reference translations. In this chapter, the authors present the results of several experiments carried out in order to check the validity of QE to: a) estimate PE effort; b) select the best translation from different MT systems outputs; c) decide if the MT output can be used for systems self-learning; and d) choose samples of translations to be evaluated manually.

The following chapter discusses the use of MT and self-post-editing as aids for academic writing and the impact that this may have on the quality of such production. The authors suggest that MT might be a useful tool to help non-native English academics achieve better quality in their writing. Some studies have explored the use of MT for second-language writing (García and Pena 2011; Sangmin-Michelle 2019). Moreover, other researchers have examined the use of lexicographical tools to help non-English speaking academics write papers in English (Laso and John 2017). However, little has been said previously about the use of MT as a tool to assist academics in their writing of papers, which makes this contribution particularly stimulating and novel.

In the last chapter of the book, Toral and Way study the level of quality that NMT can achieve when translating literature. The experiment analyses the automatic translation of twelve English novels into Catalan using an NMT system and a Statistical Phrase-Based MT (PBSMT) system. The evaluation was performed automatically (i.e., using BLEU) and manually (i.e., ranking of the PBSMT and NMT output segments and the reference translations as well). In both evaluations, the NMT output proved to be of better quality than the PBSMT one. Moreover, in human evaluation, for two out of the three novels evaluated, the judges perceived NMT output to be of similar quality to manual references in one third of the translated segments. Needless to say, this represents a significant change in the area. Two decades ago nobody would ever have considered using MT to translate literature, but the impressive performance of NMT will definitely open up new horizons in this field.

The two volumes reviewed here are complementary, in the sense that they provide a complete and accurate picture of the field of MT, covering its origins, the different types of MT technologies available and the large and complex subfield of TQA.

The first volume reviewed explores MT in a radically different way to how most traditional MT books and papers do. In an era when MT is directly linked to the use of large amounts of data and complex calculations, Scott's (2018) cognitive approach is refreshing and invites reflection. The psycholinguistic and neurolinguistic perspectives are present throughout his book, but especially in chapters three, four, five and eight, where the author seeks to understand how the brain processes natural language so that it can be simulated by MT engines. The cognitive approach is also present in Morkens et al. (2018), although to a significantly smaller degree. That said, in such a recent field as AVT, several studies related to neurolinguistics are being performed in order to better understand how the audience processes subtitles and captions. This information will be vital to improving how this type of translation is assessed.

The two volumes examined provide different but complementary views on what determines translation quality. Scott (2018) believes MT quality depends on the quality of the MT system used and the linguistic quality of the MT output. He highlights ambiguity and complexity as the two main issues that have a direct effect on quality, and he suggests ways to overcome them and thus improve MT systems performance. In his book, the quality of the MT output is specifically addressed in chapter seven, where he refers to the issues that some MT engines have developed so as to move away from source-syntax restricted output sentences. Other typical MT errors are also covered by the examples provided to illustrate the performance of MT systems when translating from English to German or French. However, translation quality is not only determined by the linguistic quality of the output; as we learn in Morkens et al. (2018), translation quality includes the above and goes beyond it. Several contributions put forward the idea that translation quality is highly dependent on the use and purpose of the translation, thus, these parameters should be taken into account in the process of evaluation. TQA must consider the perishability of the translation (e.g., reviews on the Internet), whether the translation is legally binding, if the user only wants to get the gist of the text or if the translation is constrained by certain factors (e.g., maximum length of subtitles in AVT). All the above and the fact that TQA is highly dependent on the translation provider results in a lack of standardisation, which is one of the main issues in TQA. This is another recurrent topic in the second volume. Almost all authors address the urgent need for a more standardised way to assess translation quality and agree that the goal of the translation is key.

Another common topic in both volumes is the limits of MT. In his book, Scott (2018) explains that MT is particularly limited by complexity and claims that machines are unlikely to ever translate certain types of texts, such as those where matters of style are essential (i.e., literature), as successfully as humans do. In the same line, in the second book under review, Way also suggests using raw MT for highly perishable texts and emphasises that other types of texts will definitely need MT output to be post-edited. However, the rise of new MT technologies such as NMT engines is continually pushing MT limits. The quality that NMT engines seem to be capable of achieving has shown that the use of MT in certain areas should be reconsidered. That is the case illustrated by Toral and Way's chapter, where the authors report on experiments performed using MT to translate literature. Although there is still a long way to go, we must accept that certain MT uses that were unthinkable in the past may well need to be revised in the future.

NMT systems and their performance are also discussed in both books. The first volume focuses on the internal workings of NMT systems and analyses some of their strengths and weaknesses by providing numerous examples of NMT output and comparing them to other types of engines. Morkens et al. (2018) analyses NMT's performance and reports on experiments carried out to test its quality. Both books highlight NMT's accomplishments and their positive impact in the MT area. As regards NMT's strengths, both books claim that NMT produces fewer morphological and lexical errors and substantially fewer word-order errors than other MT technologies (e.g., SMT systems). On the other hand, when analysing its weaknesses, we find out that NMT's performance degrades with sentence length. Also, the reader should note that NMT systems require large amounts of bilingual data to be trained on, which might pose a problem when dealing with less-resourced languages. I would also like to add that there is a key point about NMT's performance that I have not seen in either book and that, in my view, is vital when assessing MT quality. NMT errors tend to be semantically-based (e.g., adding words that do not appear in the source text, omitting source words or making up morphologically-similar words as explained in Forcada 2017). These errors are difficult to spot, especially as a consequence of the extremely fluent and natural NMT translated segments, which makes the process of PE particularly challenging.

Finally, I would like to draw attention to the fact that both volumes advocate for the use of linguistic knowledge in both the development of MT systems and in TQA. Nowadays most of the state-of-the-art MT systems have moved away from rules; however, Scott (2018) considers that only with the combination of the three technologies (i.e., RBMT, SMT and NMT) can developers make a significant breakthrough in MT. As regards TQA, although manual evaluation is regarded as expensive, slow and subjective, most of the authors in Morkens et al. (2018) recognize that manual evaluation is vital to improving MT systems. Thus, it should be acknowledged that despite the undeniable success of neural networks and deep-learning, linguistic information and trained linguists still play a key role in the area and should not be disregarded.

Works Cited

Bassnett, Susan. 1980. *Translation Studies*. London and New York: Routledge.

Bonhage, Corinna E. et al. 2015. "Combined Eye Tracking and fMRI Reveals Neural Basis of Linguistic Predictions during Sentence Comprehension." *Cortex* 68: 33-47.

Calzolari, Nicoletta et al., eds. 2014. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*. Reykjavik: ELRA.

Díaz Cintas, Jorge and Rocío Baños Piñero, eds. 2015. *Audiovisual Translation in a Global Context*. London: Palgrave Macmillan.

Duff, Melissa C. and Sarah Brown-Schmidt. 2012. "The Hippocampus and the Flexible Use and Processing of Language." *Frontiers in Human Neuroscience* 6: 1-11.

El-Kahlout, Ilknur Durgar et al. eds. 2015. *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*. Antalya: ACL Anthology.

Fishel, Mark et al. 2011. "Automatic Translation Error Analysis." In Habernal and Matoušek 2011, 72-77.

Forcada, Mikel L. 2017. "Making Sense of Neural Machine Translation." *Translation Spaces* 6 (2): 291-309.

García, Ignacio and María Isabel Pena. 2011. "Machine Translation-Assisted Language Learning: Writing for Beginners." *Computer Assisted Language Learning* 24 (5): 471-87.

Goto, Shinsuke, Donghui Lin and Ishida Toru. 2014. "Crowdsourcing for Evaluating Machine Translation Quality." In Calzolari et al. 2014, 3456-63.

Habernal, Ivan and Václav Matoušek, eds. 2011. *Proceedings of the 14th International Conference on Text, Speech and Dialogue*. Pilsen: Springer.

Huertas Barros, Elsa and Juliet Vine. 2017. "Current Trends on MA Translation Courses in the UK: Changing Assessment Practices on Core Translation Modules." *The Interpreter and Translator Trainer* 12 (1): 5-24.

Hutchins, John W. and Harold L. Somers. 1992. *An Introduction to Machine Translation*. London: Academic Press.

Isabelle, Pierre, Eugene Charniak and Dekang Lin, eds. 2002. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia: ACL Anthology.

Jiménez-Crespo, Miguel A. 2017. *Crowdsourcing and Collaborative Translations: Expanding the Limits of Translation Studies*. Amsterdam and Philadelphia: John Benjamins.

Koehn, Philipp. 2009. *Statistical Machine Translation*. Cambridge: Cambridge UP.

—. 2020. *Neural Machine Translation*. Cambridge: Cambridge UP.

Koehn, Philipp and Rebecca Knowles. 2017. "Six Challenges for Neural Machine Translation." In Luong et al. 2017, 28-39.

Kumaran, Dharshan, Demis Hassabis and James L. McClelland. 2016. "What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated." *Trends in Cognitive Sciences* 20 (7): 512-34.

Kurczek, Jake, Sarah Brown-Schmidt and Melisa Duff. 2013. "Hippocampal Contributions to Language: Evidence of Referential Processing Deficits in Amnesia." *Journal of Experimental Psychology: General* 142 (4): 1346-54.

Laso, Natalia Judith and Suganthi John. 2017. "The Pedagogical Benefits of a Lexical Database (SciE-Lex) to Assist the Production of Publishable Biomedical Texts by EAL Writers." *Iberica* 33: 147-72.

Luong, Thang et al., eds. 2017. *Proceedings of the First Workshop on Neural Machine Translation*. Vancouver: ACL Anthology.

Mikul, Chris. 2014. "Caption Quality: International Approaches to Standards and Measurements." *Media Access Australia*, n.d. [Accessed January 28, 2022].

Opitz, Bertram and Angela D. Frederici. 2007. "Neural Basis of Processing Sequential and Hierarchical Syntactic Structures." *Human Brain Mapping* 28 (7): 585-92.

Papineni, Kishore et al. 2002. "BLEU: A Method for Automatic Evaluation of Machine Translation." In Isabelle, Charniak and Lin 2002, 311-18.

Plass, Jan L. et al. 2010. *Cognitive Load Theory*. New York: Cambridge UP.

Popović, Maja et al. 2015. "Poor Man's Lemmatisation for Automatic Error Classification." In El-Kahlout et al. 2015, 105-12.

Romero-Fresco, Pablo and Juan M. Perez. 2015. "Accuracy Rate in Live Subtitling: The NER Model." In Díaz Cintas and Baños Piñero 2015, 28-50.

Sangmin-Michelle, Lee. 2019. "The Impact of Using Machine Translation on EFL Students' Writing." *Computer Assisted Language Learning* 33 (3): 157-75.

Zeman, Daniel et al. 2011. "Addicter: What Is Wrong with my Translations?" *Prague Bulletin of Mathematical Linguistics* 96: 79-88.

Zhang, Yaxu, Jing Yu and Julie E. Boland. 2010. "Semantics does not Need a Processing License from Syntax in Reading Chinese." *Journal of Experimental Psychology. Learning, Memory and Cognition* 36 (3): 765-81.

Elisabet Comelles Pujadas is a lecturer in the Department of Modern Languages and Literatures and of English Studies at the University of Barcelona, where she teaches English Lexicology and Morphology and Descriptive Grammar of English. She holds a PhD in Cognitive Science and Language (UB). Her research interests include Natural Language Processing, Machine Translation, Machine Translation Evaluation and Corpus Linguistics.