Deliverable Number: 3.6

Deliverable Title: Report of the state of the art on tools in computational linguistics and computer-aided translation for enhancing survey translation

Work Package 3: Maximising Equivalence through Translation

Deliverable type: Report
Dissemination status: Public

Submitted by: European Social Survey ERIC (Universitat Pompeu Fabra)

Authors: Diana Zavala-Rojas

Date Submitted: January 2018

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654221.

www.seriss.eu @SERISS_EU

SERISS (Synergies for Europe's Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe's social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe's grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey of Health Ageing and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: Zavala-Rojas, Diana (2017) *Report of the state of the art on tools in computational linguistics and computer-aided translation for enhancing survey translation*. Deliverable 3.6 of the SERISS project funded under the *European Union's Horizon 2020 research and innovation programme* GA No: 654221. Available at: www.seriss.eu/resources/deliverables

Report of the state of the art on tools in computational linguistics and computer-aided translation for enhancing survey translation

Diana Zavala-Rojas
European Social Survey ERIC

Summary

This report corresponds to the first part of Task 3.2 which aimed at investigating the suitability of tools used in computational linguistics (CL) and linguistic corpora to detect deviations in content and measurement characteristics of survey questions. In this report, we contextualise the state of the art of tools developed in the field of *computational linguistics* that can have potentially useful applications in *survey translation* i.e. the translation and translation assessment processes of survey questionnaires. We briefly review the main tools (applications and methods) in CL, thereafter, we focus on one of its prominent applications: machine translation (MT), and on the most common translation technologies that have contributed largely to the development of MT. Then, we contextualise such applications in survey translation, pointing out areas where windows of opportunity exist to incorporate tools that are commonly used in other areas of translation, and that can potentially benefit the translation and translation assessment of survey questionnaires.

1. Introduction

This report corresponds to the first part of Task 3.2 which aimed at investigating the suitability of tools used in computational linguistics (CL) and linguistic corpora to detect deviations in content and measurement characteristics of survey questions in a large scale. In this report, we contextualise the state of the art of tools developed in the field of *CL* that potentially have useful applications in *survey translation* i.e. the translation and translation assessment processes of survey questionnaires. After a brief review of the main tools (applications and methods) in CL, we focus on one of its prominent applications: machine translation (MT), and on the most common translation technologies that have contribute largely to the development of MT.

In this report we focus on the requirements that have increased the precision of MT and on the methods of MT itself, contextualised for the survey context. This application of CL was chosen following three steps outlined in the SERISS WP 3.2 task description: after an initial literature review of existing CL tools, we focused on the developments in computational linguistics relevant for enhancing both, the translation of survey questionnaires and the assessment of such translations, hence *survey translation*. From this initial literature review we learned that although applications in CL cover more developments than MT, such as speech recognition or sentiment analysis, the know-how and lessons learned gained through the development of MT can be directly useful for the development of survey translation. Moreover, during the initial literature review there were no other developments in CL with straightforward applications into survey translation.

In a second step, Task 3.2 leader and author of this report organised a first symposium in Barcelona (June 1-2, 2017) where members of four SERISS infrastructures, the European Social Survey (ESS), the European Values Study (EVS), the Survey of Health, Ageing and Retirement in Europe (SHARE) and the Consortium of European Social Science Data Archives (CESSDA) came together with experts in survey translation and CL to discuss those candidate tools that could potentially be used for survey translation. In this symposium, two CL applications were presented: MT (which will be covered in detail in this report) and the Language Effects in Surveys (LES) project (LES, 2017). In the LES project, researchers use text summarisation to identify semantic similarities and differences between words and to explore contemporary applications as well as popular usage of words across and between different languages and online populations. Although the LES project has an exciting potential to analyse the effect of wording vicinity, it is based on the analysis of millions of texts from the world wide web, that are not necessarily validated translations. Therefore, in its current stage, its application into the area of survey translation is not straightforward.

In a third step, more desk work was conducted after the symposium for completing the literature review of the history of MT relevant for the survey translation context. A key result from the literature review and the discussion session in the symposium is that outstanding developments in MT are implemented as part of *translation technologies* i.e. computer-aided translation (CAT) tools, and are used as such by language services providers (LSP), e.g. translators and translating agencies, to facilitate the translation and assessment enterprise. CAT is a form of language translation in which a human translator uses computer software to support and manage the translation process (Garcia, 2007, 2012; O'Hagan, 2009; Pym, 2011; Taravella & Villeneuve, 2013). An important lesson learned is that CAT/translation technologies have facilitated the management, storage and validation of large databases of texts that are used to train MT systems, making the history of MT and CAT tools highly interrelated and interdependent.

CAT tools are commonly used in practically all modern-day professional translation, making human translation without the aid of translation technologies very uncommon (Sin-wai, 2017). However, in the translation of academic social surveys, they are not widely used yet. An exception is the development of the Translation and Management Tool (TMT) that has been used by the SHARE. Under the SERISS project, this tool is being adapted to the translation process in the ESS and the EVS (Martens, 2017).

Therefore, by reviewing CAT tools and MT, we provide survey researchers and practitioners with valuable know-how gained in the years of development of the translation and localization industry. This allows learning about the potential advantages and challenges of a transition to more technologically oriented survey translation and translation assessment processes. For instance, the utilisation of tools in CL requires a series of prerequisites, among others, that there exists a database of texts in machine-readable format or that translations are produced, stored and managed by software. Therefore, in this report we also point out those translation technologies that had made the recent success of MT possible.

The literature on MT and CAT tools is abundant. Excellent manuscripts such as “The Future of Translation Technology” (Sin-wai, 2017) or “The Handbook of Computational Linguistics and Natural Language Processing” (Clark, Fox, & Lappin, 2013) are basic references. Throughout this report, we mention several specific tools as examples, but we do not necessarily imply that they are the best available tools, but they are mentioned as an illustration. A survey of available market and academic tools goes beyond the scope of the present research. For instance, only in 2014, 143 machine translation systems from 23 institutions were submitted to the ten translation directions in the standard translation task of the ninth workshop in MT (WMT14) organized by the Association of Computational Linguistics (Bojar et al., 2014). Thus, the aim of this report (D3.6) is to provide the basic knowledge about existing tools in MT, and those translation technologies that have enhanced it, and that can potentially facilitate the translation and assessment process of survey questionnaires[1].

Next steps in Task 3.2 are to establish a roadmap for the potential transition of survey translation into a more technologically-oriented process (D3.8). For that roadmap, a second seminar will be organised in March 2018 by the ESS at Universitat Pompeu Fabra. Establishing a roadmap for the feasibility of MT and translation technologies for survey translation should start with a systematic assessment of how the adoption of translation technologies would affect the quality of the translation output, and the translation process and assessment workflow, as it was the case in other areas of translation sciences. This report aims to be a first step in such an assessment serving as a bridge between survey translation, MT and CAT technologies to find a roadmap of potential synergies.

This report is organised as follows: In the next section after this introduction, we provide a summary of CL, as a discipline in general, thereafter we review the most common translation technologies in the translation and localisation industries. Then we highlight the areas where windows of opportunity exist to apply the translation technologies into the area of survey translation, pointing out the requirements of such a transition. Finally, we end the report with a conclusions section.

2. What is computational linguistics?

Computational linguistics is a broad interdisciplinary research field which applies statistical modelling and computational methods to language processing (ACL, 2017). Throughout its

[1] This literature review builds heavily on the content of the experts’ presentations in the “1st SERISS Symposium on synergies between survey translation and developments in translation sciences” held in Barcelona, 1st and 2nd of June, 2017 and chaired by the author of this report. Appendix 1 provides a summary of the experts’ presentations given at the Symposium.

history, research in CL has drawn upon theories in several disciplines (mainly, linguistics, psycholinguistics, mathematics, statistics, and computer science) with the main objective of developing applications derived from language modelling, for instance MT, or voice-based assistants. Theoretically-driven research in CL has been less extended, focussing mainly on syntax representation and analysis, and parsing linguistic structures (Clark et al., 2013).

The most prominent areas of application of methods in CL are: text retrieval on some desired topic; MT; question answering, from simple factual questions to ones requiring inference, descriptive or discursive answers; voice recognition; text summarization; analysis of texts or spoken language for topic, sentiment, or other attributes; dialogue agents for accomplishing particular tasks e.g. trip planning or medical advising; and ultimately, creation of computational systems with human-like competency in written and spoken dialogue (Schubert, 2015).

Methods in CL can be classified in three broad areas (Bahdanau, Cho, & Bengio, 2014; Clark et al., 2013; De Mulder, Bethard, & Moens, 2015; Fisher, 1993; Slocum, 1986; Sutskever, Vinyals, & Le, 2014): the first is syntactic-semantic frameworks i.e. methods that aimed to encode linguistic knowledge using systematic re-usable classifications. Examples of applications are ruled-based MT (Section 5 reviews MT in detail) and chatbots with numerous scripts to produce prototypical reactions to human queries. The second area of methods is about corpus-based statistical approaches; a very important paradigm which started in the late 1980's and consolidated throughout the 1990's. Statistical methods use text corpora as data to model the probability distribution of linguistic units e.g. words, sentences, paragraphs, documents, etcetera. Prominent examples are statistical MT, or systems based on text to answer questions. Corpus-based statistical methods require large amounts of machine-readable text (or speech data), a detailed study of the distributional properties of language, and the development of powerful statistically based learning techniques e.g. Bayesian estimation methods.

Finally, the third paradigm consolidated in 2014 –and which is the state-of-the-art at present- with the adoption of deep learning techniques for language modelling, i.e. algorithms that do not conduct a programmed task but that learn data representations. The main deep learning technique applied to language modelling are *artificial neural networks*, commonly referred to only as 'neural networks'. Neural networks are computing systems that imitate brain biology. They learn data structures or representations from a (very large) training dataset (De Mulder et al., 2015). They consist of several layers of information, each of them updated with new data to predict the next layer. In MT, for instance, models consist of an encoder that extracts a fixed-length representation from a variable-length input sentence, and a decoder that generates a correct translation from this representation (Bahdanau et al., 2014; Cho et al., 2014; Kalchbrenner & Blunsom, 2013; Sutskever et al., 2014).

3. From computational linguistics to computer-assisted translation

During the last two decades two interrelated fields have developed enormously, CL and CAT. These two fields are highly interrelated because scientific developments in CL are often transformed or included into CAT products. Furthermore, databases associated with CAT technologies are an important data input in CL models, for instance the Deepl MT system (Deepl, 2017) using neural networks was trained with Linguee, a huge database of bilingual sentence pairs translated by humans (Linguee, 2017).

Hutchins and Somers (1992) proposed an analytical framework helpful for explaining the interdependence of MT and CAT tools. At one end of the scale, we find human translation, which is carried out without digital aid. At the other end of the scale, *fully automatic, high-quality translation (FAHQT)* is performed entirely by a computer. Between the two extremes of the scale, we find CAT (and CAT tools). They encompass different degrees of automation, are used at different stages in the translation process, require different levels of users' knowledge about computing, and are applied to different dimensions of translation (Alcina, 2008, p. 4). In general, CAT includes human-aided MT (HAMT), in which the translation task is carried out by a computer and humans carry out pre-editing i.e. source text preparation before the translation task, post-editing i.e. editing MT output to achieve high quality translations, and disambiguation tasks. It also includes machine-aided human translation (MAHT). In MAHT the translation task is carried out by humans and computers provide different tools and functionalities to facilitate the translation task (Boitet, 1997; Varile & Zampolli, 1997). An example of HAMT is post-editing of MT output. MAHT refers, for instance, to the use of terms databases (See section 4).

4. Review of computer-aided translation tools

In the translation sciences, there is consensus about the complexity of categorising CAT technologies (Bowker & Fisher, 2010, p. 60; Folaron, 2010, p. 433; Taravella & Villeneuve, 2013, p. 65). For the public at large, machine translation (MT), is the best-known example of translation technology, driven largely by the success of services such as AltaVista's Babelfish, and Google Translate. For LSPs, mainstream CAT technologies do not necessarily include MT yet, but MAHT developments.

CAT technologies include different types of computerized tools used in the translation process (Bowker & Pearson, 2002, pp. 5–9; Sin-wai, 2017, p. 9). Although broader initial definitions cover general tools used in computing, such as word processors, e.g. Microsoft Word, OpenOffice, currently, broadly accepted definitions refer to the specific tools used in translating, such as translation memories (TM), term databases, translation glossaries and dictionaries, e.g. ProZ.com, Linguee, translation editors, corpus-analysis tools, e.g. AntConc Concordancer, and integrated management systems of translation processes and terminology, e.g. SDL Trados, Déjà Vu (all these tools are defined below).

The first CAT systems emerged in the late 1970s and early 1980s in which TM are one of the first consolidated translation technologies. TM is still the most frequently used CAT tool in professional translation (e.g. Lagoudaki, 2006; ProZ.com, 2011), alongside with databases for terms' search, controlled language tools, and translator editors. A TM is a database of validated translated segments and aligned source and target texts that translators can access to re-using previous translations. Thus, while translating, the translator is provided with translation proposals, the so-called matches, which they can accept, reject or revise (Sin-wai, 2017). It includes relevant metadata and is accessible as a bilingual, parallel corpus or interactively from the translation editor. The effectiveness of TM can be enhanced via rule-based writing of text and adhering to criteria for formatting and layout to support correct segmentation in the translation editor. This is important as a TM is only as good as its content. If a stored translation is of poor quality, this translation will be possibly accessed, and its poor quality will be spread. Ideally, TM would re-use only high-quality existing translations. However, this requires investment in quality and in meta-data on quality (Badia, 2017; Behr, 2017; Burchardt et al., 2014).

Closely related to a TM is the concept of (text) corpus. It is a large and structured database of texts, in which linguistic units are frequently parsed and annotated with respect to linguistic characteristics. Texts are subdivided into compositional and sequential units which are linked and can thus be retrieved as pairs (or triplets, etcetera). A corpus is made up of texts in machine readable format, therefore, they are the basic input to conduct linguistic quantitative analysis, such as statistical analysis, rules validation, or checking occurrences. Commonly a corpus is a body of texts in a specific domain, e.g. Shakespeare works, but it can also be a general compilation of a variety of texts that give account of how language is used in real situations, such as the Oxford English Corpus, the largest available corpus, that includes text spreading from web pages to literary works in English language (Oxford University Press, 2017).

Bi- or multi-lingual parallel corpora refer to sets of texts that are aligned and annotated, e.g., Europarl parallel corpora (Koehn, 2005), EUR-Lex Corpus (Baisa, Michelfeit, Medved, & Jakubícek, 2016), and Acquis Communautaire (Steinberger et al., 2006). Parallel corpora can also be regarded as comparable corpora on the bases of assumed similarity, i.e. corpora created according to similar design criteria that may have varying degrees of similarity. Sometimes parallel or comparable corpora include human translations, such as the Linguee database (Linguee, 2017), but this is not necessarily the case. It may contain aligned and annotated text in several languages, that are not validated translations, but monolingual texts on a same subject (Fantinuoli & Zanettin, 2015). The enrichment of a corpus with linguistic and extra-linguistic annotations plays a decisive part in language technologies used in translation. The difference between a TM and a corpus is that the TM can be considered a user-interface implementation of a corpus, with functionalities to extract text, store, edit and reuse texts and/or translations, whereas a corpus is a structured database of such texts. A TM can contain a corpus.

A second component commonly integrated into CAT technology is a database for terms search, also referred as a translation glossary, lexicon or terminology collection. A term database consists of terms and other lexical items and stores relevant information (e.g., term and equivalent(s), definition and example(s), source) (Nielsen, 2010). Similar to the first TM, it is accessible as a bi-/multi-lingual dictionary or interactively from the translation editor (Badia, 2017). Related to a terms search database are controlled vocabulary tools, such as thesauri; they are lists of words of terms grouped in synonyms or relating concepts (Aitchison, Gilchrist, & Bawden, 2000). The construction of domain specific thesauri supports the translation and assessment tasks by providing disambiguation or giving examples of language use; thesauri are also broadly used for indexing archives. For the Social Sciences, there is the European Language Social Science Thesaurus (ELSST) (Balkan, Jääskeläinen, Frentzou, & Kappi, 2011), a multilingual thesaurus developed by the Consortium of European Social Science Data Archives (CESSDA) that is utilised for indexing social science surveys (See SERISS Task 3.3).

A third broadly accepted component in CAT technology is a translation editor, which interacts with the TM and the term database. A common task in a translation editor is to pre-translate the text of interest, and validate the final translation (Carson-Berndsen, Somers, Vogel, & Way, 2009; Nielsen, 2010; Somers, 2003). Although these three tools suggest translation proposals, they should not be confused with MT which will always involve an algorithm to produce translations. Examples of commercial integrated translation technologies that include TM, term databases and translation editors are Systran (which also provides machine translation) (SYSTRAN, 2017) and SDL Trados (SDL PLC, 2017). Open source translation technologies also exist such as OmegaT (OmegaT, 2017) or Anaphraseus (Anaphraseus, 2017), for creating translation memories; or integrated translation software combining management and translation functionalities such as OmegaT+ and Open TM2 (OpenTM2 Steering Committee, 2017).

In addition to the three basic components of CAT technology mentioned above, LSPs have broadly accepted systems for automated checks e.g. RegExr (Rinker, 2014) and regular expressions. Automated checks compare source and target text with respect to different characteristics, such as punctuation, length, numbers and dates; regular expressions, i.e. regularly used terms and expressions, are stored in a database which are verified with automatized procedures (Behr, 2017; Souto, Lupsa, & Dept, 2017).

Acceptance is also large for management systems of terminology and of the translation process itself; for instance, in the translation of social surveys the TMT can be categorised as a program for management of the translation process. The adoption of CAT technologies for process management has facilitated translation done through teamwork, as translation professionals are commonly linked by a server-based CAT system that manages the workflow and sometimes invoicing and accounting of professionals involved in a project. However, it has also influenced enormously the translators' work, as they must be familiar with different software functionalities for a variety of tasks. The CAT tools databases have derived into training data for state-of-the-art MT systems. For instance, Deepl.com, which was regarded as the best automated translation system in 2017 according to blind tests comparing the MT output with the translations produced by a human, is trained with the Linguee multilingual corpora.

Corpus-analysis tools, e.g. AntConc Concordancer, are a large body of computer programs that process and analyse monolingual and multilingual corpora, for applications in translation and in CL in general (Berberich & Kleiber, 2017). Some examples of the most common tools are software to annotate texts such as @nnotate, WebLicht; concordancers create lists of main words in a body of texts; file converters converting pdf files into plain text files, for instance AntFileConverter. There are tools for different types of text analysis, e.g. to check for occurrences (CorpusTools), vocabulary complexity (Online Graded Text Editor), or tools for corpora creation, such as textSTAT, among others.

Sin-wai (2017) estimates that over 90 per cent of all translation projects on the global translation market are from the area of 'practical translation' (that is, technical translations, focussing on specific domains, such as business, economics, legal, medical, technology etc.) and only 10 per cent is about literary translation. Practical translation includes the translation of survey questionnaires. In most practical translation, translators reuse and recycle previously translated segments, as the original texts are often repetitive. The adoption of the aforementioned CAT technologies by LSPs facilitates largely repetitive practical translation by potentially eliminating duplicate work (Merkel, 1993). However, adopting CAT technologies also brought changes in the translation process and the adoption of new steps in the workflow. For instance, pre-editing texts was soon seen as a key step for translators aiding their work with CAT software, as CAT systems only process texts without format, e.g. files with *.txt extensions (Sin-wai 2017). Assessing the quality of translations was also of key importance in the transition to technologically-oriented professional translation. A precondition to reuse translated texts, terms and expressions stored in databases and in translation memories is that texts have to be validated with respect to their quality (Fantinuoli & Zanettin, 2015; Murray, Dorr, Lin, Hajič, & Pecina, 2006). Versioning has also been of key interest in translation technologies. Draft translation outputs are preserved, under the assumption that intermediate drafts before the final translation are meaningful to understand the translation process and the complexity of the translation task.

5. What is machine translation?

MT is a highly multidisciplinary subfield of CL whereby human translators, engineers, computer scientists, mathematicians, statisticians, and linguists share research objectives. Sin-wai (2017) regards machine translation as pre-translation on two counts. First, it involves preparatory work on the texts to be translated, including checking the spelling, compiling dictionaries, and adjusting the text format. Second, it is taken to be a draft translation of the source text, which can be further revised by a human translator.

A widespread criterion to classify MT approaches includes the sources of information they use to produce a translation. These could be rules, data, or hybrid systems (Costa-Jussa et al., 2016; Costa-Jussa & Fonollosa, 2015). MT approaches based on rules (RBMT) use linguistic information, e.g. dictionaries, in combination with linguistic knowledge in three different phases: analysis, transfer, and generation to transform a source language text into a target language text e.g. Systran, Apertium (Forcada, Tyers, & Ramírez Sánchez, 2009). In RBMT a set of rules for a given language combination are programmed to produce a translation.

Data-driven MT -statistical machine translation (SMT) and neural machine translation (NMT)- is prevalent within the MT academic research community (Costa-Jussà & Farrús, 2014). Data-driven MT approaches use information from data, mainly parallel corpora, and algorithms to model translation.

Until recently, data driven MT was statistically-based, at present it also refers to NMT. SMT e.g. Moses SMT system (Koehn et al., 2007), is probability-based and relies on large corpora of source texts and their human translations (Brown, Pietra, Pietra, & Mercer, 1993). It uses probabilistic models to produce translations. SMT approaches learn from training data, mainly parallel corpora, aligned at the sentence level. Given a source language string of text, the goal is to choose the string with the highest probability among all possible strings in the target language (Costa-Jussà & Farrús, 2014).

Before the appearance of NMT, research had focused on narrowing the boundaries between rule-based and statistical MT approaches, by investigating hybrid MT systems, a combination of RBMT and SMT, that would take both programmed rules and data to produce texts in the target language (Eisele et al., 2008; España Bonet, Màrquez Villodre, Labaka, Díaz de Ilarraza Sánchez, & Sarasola Gabiola, 2011).

NMT is a newly emerging approach to MT. Examples of NMT programs are Google Translate (Sutskever et al., 2014), Microsoft Translator (Microsoft Corporation, 2017), Deepl (Deepl, 2017), and OpenNMT (Klein, Kim, Deng, Senellart, & Rush, 2017). Bengio et al., (2003) proposed a model that would learn simultaneously a representation for each word of a training dataset along with the probability function for word sequences, expressed in terms of these representations. From this work, NMT models developed that belong to a family of encoder–decoder models (Bahdanau et al., 2014; Bengio, Ducharme, Vincent, & Jauvin, 2003; Cho et al., 2014; Sutskever et al., 2014). An encoder neural network reads and encodes a source sentence into a fixed-length vector. A decoder then outputs a translation from the encoded vector. The encoder–decoder system consists of the encoder and the decoder for each of the languages which are jointly trained to maximize the probability of a correct translation for a source sentence.

Traditional SMT systems (see e.g. Koehn et al., 2007) consist of many small sub-components that are tuned separately, while NMT (Kalchbrenner and Blunsom 2013), Sutskever et al. (2014) and Cho et al. 2014b) builds and trains a single, large neural network in a language pair jointly that reads a sentence and outputs a translation. NMT has improved considerably the quality of MT output, such that, historical SMT systems are transitioning from probabilistic models to neural models. Examples are the open source MOSES SMT system, and eTranslation (Directorate-General for Translation, 2017), a Connecting Europe Facility (CEF) building block providing multilingual MT services for European and national public administrations.

Until recently the LSP community's acceptance of MT was considerably small; nowadays, however, the acceptance is becoming larger, especially when MT is used in combination with human translation and post-editing, i.e. the process by which humans fix machine translation outputs to achieve a high quality final output (Claverie, 2014). Although post-editing MT outputs is becoming a popular procedure, Burchardt et al. (2014, pp. 34–35) argue that in order for LSPs to embrace MT, they need automatic tools to recognize MT output in at least three categories: 1.) MT output that can be used as is. 2.) MT output that can easily be post-edited to meet specifications. 3.) MT output that should be discarded. They argue that when tools identify these quality categories, rather than providing generic metrics for large amounts of translated text or expecting humans to post-edit all content, then the tools can work with humans to leverage their strengths.

The quality of the data-driven MT output is highly dependent on the quality of the training data; therefore, the creation of parallel corpora has been a very important factor for the development of MT. Pre-editing of corpora data i.e. the preparatory work of alignment, annotation and format of the source texts, is an area of extensive research (Fantinuoli et al., 2015). Currently, efforts are being placed into training NMT systems using domain-specific monolingual or multilingual parallel corpora (see e.g. Directorate-General for Translation, 2017).

6. Translation process in comparative surveys, opportunities and challenges for CAT technologies

In the same two decades of rapid development of CAT and CL, multinational, multiregional and multicultural (3MC) surveys have consolidated as important sources of data to compare social and political phenomena across cultures. With the 3MC surveys landmark came the increasing significance of rigorous multilingual translation and assessment of survey questionnaires, as low quality translations hamper data comparability (Mohler, Dorer, De Jong, & Hu, 2016). Survey translation has developed best practice procedures that bring together practices in translation sciences and the particular needs of survey research (Harkness, 2003; Harkness, Villar, & Edwards, 2010). The goal is to achieve *functionally equivalent* survey instruments in multilingual contexts, i.e. that across groups being compared, the indicators obtained from survey instruments represent the same concepts they intend to measure (Zavala-Rojas, Saris, & Gallhofer, 2018, in press) .

Harkness (2003) suggested that a good translation aiming at functional equivalence would avoid changing deliberately other semantic components than those necessary because of language differences. This does not imply that a literal, i.e. word-by-word, translation is looked for but that questions should keep the same concepts across languages, maintaining their intended psychometric properties. To achieve high-quality functionally equivalent translations, Harkness suggested the Translation, Review, Adjudication, Pretesting and Documentation (TRAPD) method, a team or committee approach in a multistep process where different members provide their specific expertise to arrive at a final translation.

The TRAPD method implies that at least two translators should produce independent and parallel translations from the source version into the respective target language ('T' in the acronym). In a joint translation team or 'review' meeting, the translations are discussed by the two translators together with the 'reviewer' ('R'), and the adjudicator ('A') is responsible for the final decisions on different translation options. The translated questionnaire is pretested before fieldwork ('P') and the whole process is documented ('D'). Ideally, the team members combine survey knowledge, translation and linguistic expertise, knowledge about the culture where it will be administered, and potentially other knowledge related to the topic of the questionnaire (European Social Survey, 2016).

Variants of the TRAPD method are used to translate questionnaires in cross-national surveys such as the ESS (European Social Survey, 2016), the EVS (Przepiórkowska & Behr, 2017), SHARE (Harkness, 2005), the Programme for International Assessment of Adult Competencies (PIAAC) (OECD, 2014), in medical and health research (Forsyth, Kudela, Levin, Lawrence, & Willis, 2007; Kietzmann, Wiehn, Kehl, Knuth, & Schmidt, 2016) and more recently the approach is being explored for translation of questionnaires in market research (Sha & Lai, 2016).

A backward of the TRAPD process is that it is very labour intensive, as it relies on multiple human translators and other experts. Related to this limitation is that MAHT and HAMT have not been extensively explored yet. Although to different degrees, the ESS, the EVS and the SHARE started to make use of translation technologies, with the Translation Management Tool (TMT) as the first joint project. Initially developed by CentERdata for SHARE (Martens, 2017), and adapted under SERISS for the ESS and the EVS. The TMT allows managing and documenting the TRAPD approach. However, in its current version, it does not incorporate other broadly used translation technologies, such as term database and translation memories, although developments into these two directions are ongoing.

In the translation and localisation industry, the emergence of translation technologies was accompanied by an assessment of how the quality of the translation and the translation workflow would be affected when translation processes changed into more technology-oriented procedures (Sin-wai, 2017, p. 45). In the years to come, the same rationale should be applied to survey translation in its potential transition to adopting CAT technologies, even if they have already been accepted for a long time in other areas of translation. This assessment can also be extended to improve the current functionalities of existing technologies such as the TMT. The broad adoption of CAT technologies required a series of prerequisites, for instance, pre-edition of the source texts is key to obtain texts in machine-readable format, aligned, and segmented, and ready to incorporate annotations and human edition (Serbina, Niemietz, & Neumann, 2015). It is also of key importance that texts are converted into databases such as multilingual corpora, allowing quantitative analysis of the translation efforts. Intuitive and user-friendly interfaces for translation memories can facilitate translators' work, but if they are not well designed they can also hamper the human translation tasks; therefore it has been of key importance that the development of tools includes the users' perspective and feedback (Pym, 2011). Finally, translation production, storage and management by software allows documenting valuable information of the translation process (Bowker & Fisher, 2010).

Badia (2017) suggested that an assessment of the transition to a more technologically oriented approach in survey translation can start by dissecting the TRAPD process and further verification steps to analyse which translation technologies can be applied at each of these steps. Although such an assessment is a large investment, in terms of research resources, survey researchers can rely on the large literature and know-how developed in other areas of translation. At present, LSPs are incorporating technological advances of recent years by combining various approaches and tools to maximise the benefits of them. Examples of such combinations include hybrid MT systems and systems that integrate TM, terminology management and MT run by human translators (Christensen & Schjoldager, 2016). MT is implemented by integrating it into a TM whereby a MT match is provided when no match can be found in the TM database (Bundgaard, Christensen, & Schjoldager, 2016; Koby, 2013).

As an example, a general questionnaire translation and assessment process with technological support may look like the following (Badia 2017): In a first step, the source text is pre-edited in a way that ensures that the text can be easily handled by the automated translation process. The pre-edited text gets then automatically translated. Finally, this translation is post-edited either manually or semi-automatically to ensure that the quality requirements are met, resulting in the target text. Oftentimes the source text is pre-edited using specific rules to control the language (e.g., write sentences that are shorter than 25 words or write sentences in active form).

Table 1 illustrates examples of translation technologies that can be integrated in the survey translation process alongside with a summary of some of the requirements to such integration, expected according to the literature and know-how from other areas of translation. In this report, these areas have been detected, further desk work and the next SERISS Task 3.2 symposium will help in further specifying the areas of opportunity.

Table 1: Summary of possible translation technologies for the survey translation process and its requirements

Survey translation stage	Example of CAT tools, methods	Some requirements
Translation, Review, Adjudication	Process management tools	– User-friendly interfaces for translation teams – Translation team training
	Translation memories	– Validated translations – Pre-edited texts in software readable formats – Standardized text formats – Search engines – Post-editing functionalities for retrieved translation output
	Translator editors, i.e. allowing user friendly text processing, automatic copy and paste of already translated texts, spelling checks, et cetera.	– Programming of text processing tools: copy and paste commands, spelling checks, grammar checks
	Terminology management tools	– Controlled language tools (Thesauri) – Validation of translations of terminology
	Machine translation	– Parallel corpora of validated translations – Meta-data annotations – Linguistic annotations – Pre-edited texts in software readable formats – Training of algorithms – Metrics of MT output quality – Functionalities of post-editing of MT output
Pretesting	Tools to enrich the translation memories and the terminology systems	– User friendly functionalities for building/improving databases, e.g. pretesting teams can provide feedback directly in the tool
	Terms database (Problematic) terms database	– User friendly functionalities for building databases – Archiving of problematic terms in the translation and assessment of social surveys
Documentation	Versioning management tools	– Defining what constitute a new version of a translated survey item and programming of markers.
Quality assurance	Systems for automated checks	– Extensive programming of automated checks. – Definition and list of characteristics that can be automatically checked
	Machine translation	– Parallel corpora of validated translations – Meta-data annotations – Linguistic annotations – Pre-edited texts in software readable formats

		– Training of algorithms – Metrics of MT output quality – Functionalities of post-editing of MT output
Semantic verification	Terminology management tools	– Controlled language tools (Thesauri) – Validation of translations of terminology
	Problematic terms database	– User friendly functionalities for building/improving databases, e.g. pretesting teams can provide feedback directly in the tool

7. **Conclusions**

This report corresponds to the first part of SERISS Task 3.2 which aimed at investigating the suitability of tools used in computational linguistics (CL) and linguistic corpora to detect deviations in content and measurement characteristics of translated survey questions. The aim of the report was to contextualise the state of the art of tools developed in the field of *computational linguistics* that can have potentially useful applications in *survey translation* i.e. the translation and translation assessment processes of survey questionnaires.

After a brief review of the main tools (applications and methods) in CL, we focused on one of its prominent applications: machine translation (MT), and on the most common translation technologies, CAT tools, that have contributed largely to the development of MT. Then, we contextualised such applications in survey translation, pointing out areas where windows of opportunity exist to incorporate tools that are commonly used in other areas of translation, and that can potentially benefit the translation and translation assessment of survey questionnaires.

Translation technologies may offer a variety of potential solutions to survey translation needs. The introduction of technology within survey translation requires an analysis of the processes, a determination of automatable actions, and a modification of human procedures.

Many translation tools are already available. Further research will be focused on the thorough assessment of TM and MT for survey translation and assessment, with a detailed analysis of their challenges and potentials. This topic will be the focus of the "2nd SERISS Symposium of the synergies of translation technologies and survey translation."

Therefore, the next step within SERISS Task 3.2 is to develop a roadmap to evaluate the extent by which these two technologies can be brought into survey translation. The first area of translation technology selected for further investigation is TMs (and parallel corpora). TMs are selected because translation tools based on TMs were the first ones commonly used by LSPs in other areas of translation and continue to be the most used ones at present. Therefore, the author suggests building on the existing know-how in other areas of translation. Validation of translations that feed the TM is a key area of research.

The second type of translation technology to be further investigated is MT. MT is chosen because it introduces high potential benefits to the survey translation and assessment process. Sin-wai (2017) argues that the quality achieved by some MT systems makes it worthwhile to be used in some areas of translation, when combined with human post-editing, but not in all areas, therefore it is worth investigating its potential and risks in the translation and assessment of survey questionnaires.

Behr (2017) suggests that for MT to eventually be used as a useful tool for certain processes in survey translation, it first requires the realization of quality-impact evaluation studies as well as the involvement of translators or post-editors in the development of this new step in the process, and targeted training of these groups. Extensive testing of new steps incorporating MT should be conducted, and therefore the design of pilot tests of new procedures is also of key importance. Risk management should accompany the use of MT to decide where its use makes sense and where it does not make sense.

In addition, MT can be an area to explore for the future development of the TMT. Currently, SERISS survey researchers and TMT developers are testing the tool's functionalities to manage and document the translation process. The TMT developers are currently working on integrating TM technologies into the TMT. So further developments can be addressed to strengthening the TMT's potential as a CAT tool by incorporating, in addition to TM technologies, linkages to MT technologies.

We acknowledge that Task 3.2 is limited to research and it does not include implementation. It is out of scope to conduct tests that implement language technologies into survey translation. However, we consider that a careful assessment can increase the chances of success of eventually implementing such tests. As in other areas of translation, a multidisciplinary collaboration of academia and the private sector is an important step to a successful transition. Therefore, Task 3.2 will continue nurturing a newly established network of experts that involve survey practitioners, LSP in survey translation, survey researchers, linguists, translators and translation scholars and computational linguists.

Appendix 1. Summary of the experts' presentations in the "1st SERISS Symposium on synergies between survey translation and developments in translation sciences" (1& 2 June , 2017, Universitat Pompeu Fabra, Barcelona)

1. Translation process in comparative surveys

Translation process in the European Social Survey, Brita Dorer

In "Questionnaire translation process in the European Social Survey (ESS)" Dorer describes the multi-step, multi-disciplinary approach used in the cross-national survey ESS to guarantee equivalence between all the survey item translations in all the participating countries. Starting point for the translation is the British English source questionnaire which is translated by each national team. The general translation process consists of three parts and is based on the TRAPD method (translation, review, adjudication, pretesting, and documentation; Harkness, 2003). The first part includes the implementation of the TRAPD, two translators, with ideally at least one professional linguist or translator, translating the questionnaire (T), both translator and the reviewer, an experienced survey researcher, reviewing and discussing the questionnaire translation (R), and an adjudicator, who is both familiar with the survey and has good translation skills, signing-off on the translations (A). In the second step, the adjudicated version is sent to an external service provider called cApStAn for semantic verification. In the third step, the formal characteristics of the translated question get coded with the freely available web program Survey Quality Predictor (SQP; Saris et al. 2011). These codes are then compared to the codes of the British English source version. Before arriving at the target instrument, the translated question is pre-tested (P), and the entire process is documented (D). With the start of ESS Round 8, the Translation Management Tool (TMT), developed at CentERdata in the Netherlands, has been used by a selection of volunteering countries to test its functionalities.

Translation process in the Survey of Health, Ageing and Retirement in Europe, Yuri Pettinicchi

In "SHARE – Translation procedures" Pettinicchi explains the translation-related procedures utilized in the multi-disciplinary and cross-national Survey of Health, Ageing and Retirement in Europe (SHARE). In particular, the author refers to the questionnaire development, the translation phase, and building the computer-assisted personal interview (CAPI) instrument at national level. SHARE utilizes the online-based Translation Management Tool (TMT) that in a latter phase is used to create the national CAPI instrument. In a first phase, the questionnaire is developed, which is coordinated internationally. An item glossary with background information for the translators and with interviewer instructions is generated. In a second phase (i.e., the translation process), SHARE differentiates between items of previous waves, that only get revised, and new items, that follow the TRAPD (translation, review, adjudication, pretesting, and documentation) translation method (Harkness 2003). In the TRAPD, two independent translators, who belong to the country teams, produce a first draft. A second draft is developed based on the feedback of experts. A reviewer (i.e., the country team operator)

then advances the optimal version, and an adjudicator (i.e., the country team leader) takes the final decision. The process is documented with the TMT.
Once the translation process has been finalised, the national CAPI instrument is built on the basis of the TMT. Then the country teams test their national CAPI instruments extensively. Once all the errors are corrected, the national CAPI instrument is ready to go on-field.
Two test-runs, named “pre-test” and “field-rehearsal”, provide additional feedback on the quality of the translation. Country teams have a debriefing session after each test run with their national survey agency.

An interesting note is that, for the seventh SHARE wave, the inclusion of eight new countries constituted a special case within the translation phase. The translation process for these new countries is similar to the one of a new item, but with the difference that the first draft is made by an external translator company (i.e., cApStAn). The decision to deviate from the standard procedure was made to speed up the inclusion of inexperienced country teams.

Translation process in the European Values Study, Ruud Luijkx and Claudia Brunori

In “Translation in the European Values Study”, Luijkx and Brunori illustrate the translation process of the repeated cross-sectional European Values Study (EVS). The current wave (2017-2018) encompasses at least 25 countries, with several multilingual countries and so-called ‘shared’ languages (fielded in more than one participating country). The translation of the master questionnaire focuses on the principles of comparability across countries, comparability across time and consistency between shared languages. The national teams translate the EVS questionnaire following a simplified version of the TRAPD method. For the new and modified items, two independent translations are generated, which are reviewed and then adjudicated. Items that were left unchanged since the previous waves are solely reviewed and adjudicated. In the case of shared languages, two alternative harmonisation strategies exist: adaptation and cooperation. In the adaptation approach, one country translates, reviews, and adjudicates the questionnaire. Later on, this translation is used as a base for the other countries’ adaptation. In the cooperation approach, the countries sharing a language work together throughout all the process. In the current fifth wave (2017), the entire process is implemented, handled and documented via the Translation Management Tool (TMT) developed by CentERdata.

2. Technological advancements in translation studies - what do they offer to survey translation?

Technological and computational advances in translation research and industry, Dorothée Behr

In “Technological and computational advances in translation research and industry”, Behr gives an overview of what translation sciences and current practice in other fields have to offer to survey translation. In particular, the author highlights two types of tools: machine translation (MT) and computer-assisted translation (CAT) tools. For MT to eventually be used as a useful tool for certain processes in survey translation, it first requires the realization of evaluation studies as well as the involvement of translators or post-editors in the development process, and targeted training of these groups. Risk management should accompany the use of MT to decide where its use makes sense and where it does not make sense. Four different MT approaches can be distinguished: (1) rule-based MT (RbMT), which involves a set of rules for a given language combination; (2) the statistical MT (SMT) approach, which is probability-based and relies on large corpora of source texts and their human translations; (3) hybrid systems, which are a combination of RbMT and SMT; and (4) neural MT (NMT), which involves deep learning.

Second, CAT tools include machine-aided human translation and human-aided machine translation. These kinds of tools generally segment text into smaller parts (e.g., into sentences) and store the translation units (i.e., the combination of the source and the target text/translation), thereby creating translation memories (TM) that can be re-used. CAT tools can integrate MT, terminology databases, and automatic or manual look-up. In addition, they may integrate automated checks that compare source and target text with respect to different characteristics e.g., punctuation, length, numbers and dates, regular expressions or RegExr (See Section 4).

The effectiveness of TM and MT can be increased via rule-based writing of text (especially for non-native speakers) and adhering to criteria for formatting and layout to support correct segmentation in CAT tools and TM leverage. This is important as a TM is only as good as its content. If a stored translation is of poor quality, this translation will be possibly accessed and its poor quality will be spread (i.e., "garbage in, garbage out" or GIGA principle). Ideally, TM would re-use good already existing translations. However, this requires investment in quality and in meta-data on quality.

Many translation tools are already available (e.g., authoring tools, controlled language tools, translation and terminology management systems, machine translation systems, quality assurance tools). Before building new tools, it is important to manage the expectations of potential users, to involve stakeholder, and to ensure the easy integration of the already existing tools.
Translators' competences should not only include linguistic, thematic, and intercultural abilities, but also IT-related ones (e.g., technological knowledge for utilizing CAT or other tools, information mining competence). Finally, the author describes three trends: community/user-generated translation, open access and sharing, and translation-oriented authoring.

3. Making use of past experience: managing the translation process in large scale surveys

NeferTT: How technology can improve quality and efficiency, Oscar Rivière.

In "NeferTT: How technology can improve quality and efficiency" Rivière describes NeferTT as an online management and translation tool that is designed to be used from the source questionnaire, through a multi-step translation process to the final scripting of local fieldwork questionnaires. The tool is currently used within Kantar Public. According to the author, NeferTT provides an environment for fast and reliable translations that may help to streamline the entire work process. It includes a database of previous projects. Its user-friendly environment is characterized by displaying all the information in one place and on one screen, by tracking all the changes that occur during the different steps of the translation process, and by offering a communication platform for coordinators, the translators and the fieldwork partners to exchange instructions, issues to follow up or any type of comment. The program makes a distinction between new items (i.e., items to be translated), recycled items (i.e., items that already exist in the database), trend items (i.e., items that have been used in previous waves of the survey and that should not be altered), and modified trend items (i.e., items that have been used in previous waves of the survey, but need to be slightly modified to account for changes in dates, places, names, etc.). Over a hundred users can be registered for one single project on NeferTT.

Asking MOSES to help with translation verification, Yuri Pettinicchi

In "Asking MOSES to help with translation verification" Yuri Pettinicchi exemplifies how MOSES, a statistical machine translation system (SMT), is used within the Survey of Health, Ageing and Retirement in Europe (SHARE) to verify item translations.

The general aim of utilizing MOSES is to prevent avoidable mistakes and to improve the quality of the data. SHARE handles translation within the Translation Management Tool (TMT) environment. Issues that may be detected by MOSES are the misspelling of words or commands, empty fields or missing sentences, and flipped translations. Currently, SHARE relies on the visual checking of the TMT and on the testing of the national version of the CAPI.

So-called sanity checks may automatically check for empty cells, duplicates, numbering, etc. These checks result in flagged items and possible feedback for the country teams: the main innovation is to use MOSES as an additional sanity check. MOSES is a system that can be trained to find the highest probability translation of a given sentence. Using MOSES for back translation, one can afterwards measure the distance between the back translation and the source sentence (both in the same language). One possible metric is the count of words that the two versions have in common. The outcome of this check can be reported to the project manager as a list of items that should be checked by the country team operator.

4. Survey translation and assessment possibilities for incorporating automated procedures

Manuel Souto, Danina Lupsa and Steve Dept

In "Applying translation/localization best practices in international surveys" Lupsa, Souto Pico and Dept give an overview of the life cycle of a language version of an international survey, the translation industry's best practice, and what cApStAn, a company specialized in linguistic quality assurance (LQA) and linguistic quality control (LQC) of survey and assessment materials, has to offer. The authors distinguish four stages within the multilingual content life cycle of international data collection instruments. In the first stage, item developers or authors create the source content. Next, localization engineers generate the translation files, which, in the third stage, are translated into the national target languages by translators. Lastly, the final target versions are subject to quality evaluation (by linguists) and to automated checks, using a variety of LQA applications. The translation industry's best practices include the optimal preparation of projects for translation, ensuring an optimal working environment and leveraging language assets, and providing linguistic quality assurance and quality control. However, these best practices (e.g., complying with standards, using already available technology) are not always followed, particularly in large survey localization projects. Instead, survey researchers oftentimes create new translation interfaces that lack even the most basic functionalities and ergonomics readily available in all existing computer-aided translation (CAT) tools. As a result, complexity increases, and time and money are unnecessarily spent. These suboptimal processes then lead to suboptimal results.

cApStAn improves the translation process and its outcome(s) through the following four means: First, cApStAn aims at optimizing the source text. This includes controlled writing, appropriate file preparation, translatability assessment, and the creation of project-specific rules. Second, the company focuses on the preparation of the translation and the adaptation process. This incorporates creating glossaries, style guides, language-specific rules, and translation and adaptation notes. This may also involve trend management (i.e., content transfers), a repository for documentation as well as training translators and verifiers. Third, the actual translation and adaptation process consists of a combination of one or several of the following: double or single translation, reconciliation, (team) adjudication, consultation with domain experts, and proofreading. Fourth, the linguistic quality process features automated checks, translation verification, review of verification feedback, post-verification final checks, reports, and updated translation memories (TM).

(Re-)translating thesaurus terms: lessons learned and prospects for automating survey translation, Lorna Balkan

In "(Re-)translating thesaurus terms: lessons learned and prospects for automating survey translation" Lorna Balkan focuses on the contrast between translating thesaurus terms and survey questions. The author defines a thesaurus as a set of terms that represent concepts in a domain, which may be single or multi-word units (e.g. health, community health services) and are related to each other either hierarchically (e.g. family members, grandchildren, grandparents) as well asor non-hierarchically (e.g. parents, parental role). Thesauri are principally used for indexing (i.e. assigning terms to data/documents) and retrieving information, since they control for synonymy and ambiguity. The European Language Social Science Thesaurus (ELSST) is a multilingual thesaurus developed by the Consortium of European Social Science Data Archives (CESSDA) and CESSDA-related archives for indexing social science surveys. It is used for cross-language information retrieval in the CESSDA data portal. The work done within the SERISS framework centres on evaluating thesaurus translations by (re-)translating thesaurus terms and comparing the ELSST terms assigned to a set of cross-national studies across CESSDA and its associated archives.

Thesaurus translation and survey translation may learn from each other and share automatic tools. They are similar with respect to the subject matter (i.e. social sciences) and overlapping terminology. They also share similar challenges related to semantic and syntactic ambiguity, cross-language equivalence (i.e. due to linguistic and cultural differences), and context dependence. However, they differ from each other in the following ways: the length of text to be translated (i.e. short vs. long), their grammatical form (i.e. term vs. sentence), their context (i.e. broader/narrower/related terms vs. preceding text in question/previous questions), and their metadata (i.e. qualifiers/scope notes vs. explanatory notes).

5. Technology in translation studies and translation assessment

Antoni Badia, UPF

In "Machine Translation and related technologies for translation" Toni Badia gives an overview of translation technologies. The author traces the basis or later development of machine translation (MT) back to the 13^{th}-17^{th} century, in which the first ideas of developing universal languages (like the ones developed by Ramon Llull, Gottfried Leibniz, and Johann Joachim Becher) were found. The first patents for mechanical translation as well as the deciphering of German secret code in World War II are seen as further predecessors of MT in the late 19^{th} and early 20^{th} century. Throughout the last 80 years MT has become increasingly important.

Three types of MT can be distinguished: rule-based MT (e.g., direct, transfer, or interlingua), data-driven MT (e.g., example-based and statistics-based MT), and neural MT (i.e., a neural network for translating). The author views statistics-based MT (SMT) as the state-of-the-art system, provided that the language pairs are reasonably similar in linguistic structure and word order and that there is a large amount of parallel data in the desired domain. Currently, neural MT is reported to achieve similar results as the probability-based SMT (PBSMT). Stable systems may be of any type, but three basic requirements need to be fulfilled: appropriate insertion in production environment, maintenance facilities, and trained users. The results of MT may be evaluated by humans (i.e., with respect to adequacy, fluency, and informativeness), or they may be evaluated automatically by drawing on distinct metrics with different attributes (e.g., correlation with human judgments, sensitivity to differences between MT systems).

As for today, computer-assisted translation (CAT) is composed of three major components. Its first component is translation memory (TM). That is a database of translated segments obtained from validated translation segments and aligned texts. It includes relevant metadata and is accessible as a bilingual, parallel corpus or interactively from the translation editor. A second component is a term database, which consists of terms and other lexical items and stores relevant information (e.g., term and equivalent(s), definition and example(s), source). Similar to the first component, it is accessible as a bi-/multi-lingual dictionary or interactively from the translation editor. A third major CAT component is the translation editor, who interacts with the TM and the term database. The translation editor may pre-translate the text of interest or validate the final translation. Of course, CAT may include other components too (e.g., MT, project management, quality assurance checks, term extractor).

A general translation process with technological support may look like the following: In a first step, the source text is pre-edited in a way that ensures that the text can be easily handled by the automated translation process. The pre-edited text gets then automatically translated. Finally, this translation is post-edited either manually or semi-automatically to ensure that the quality requirements are met, resulting in the target text. Oftentimes the source text is pre-edited using specific rules to control the language (e.g., write sentences that are shorter than 25 words, write sentences in active form).

In sum, translation technology offers a variety of solutions to translation needs. Among other things, these solutions include the creation and maintenance of a database of translated sentences as well as the creation and maintenance of a multilingual database with 'difficult' terms or other lexical elements. Another solution is the adaptation of the controlled language strategy. Lastly, the introduction of technology within the translation requires an analysis of the processes, a determination of automatable actions, and a modification of human procedures.

6. Research to assess translation effects

Comparing Public Opinion - Language Effects in Comparative Survey Research, Stefan Dahlberg.

In "Comparing public opinion language effects in comparative survey research" Stefan Dahlberg presents the Language Effects in Surveys (LES) research project, an inter-disciplinary project between computational linguistics and political science funded by the Swedish Research Council. It is a collaboration of three universities (i.e., University of Gothenburg, University of Bergen, University of Toronto), the Swedish Institute of Computer Science (SICS), and the Swedish language technology company Gavagai. This methodological project aims at exploring the meaning and usage of concepts relevant to public opinion research across a variety of languages and cultures. In particular, language effects occur when the language of administration of a study affects respondents' answers. A requisite in comparative country studies is that the measurements are not only valid and reliable, but also equivalent. However, it still remains an empirical question how one can know that semantic equivalence is achieved. To answer this question, the LES relies on formal epistemology through statistical methods by implementing survey experiments and distributional semantics. The latter method makes it possible to identify semantic similarities and differences between words and to explore contemporary applications as well as popular usage of words across and between different languages and online populations.

Striving for evidence - research on effects of different translation variants on data Dorothée Behr.

In "Striving for evidence – research on effects of different translation variants on data" Dorothée Behr presents a study on translation effects that is based on the observation that earlier rounds of the European Social Survey (ESS) show many shared-language differences. That is, a source question was translated slightly differently across countries that share the same language (i.e., German in Austria, Germany, and Switzerland). The study relied on the GESIS Online Panel Pilot, a web panel that included the German-speaking population residing in Germany of 18 years and older. 466 respondents of this panel took part in the survey and received a split-ballot experiment. The items tested came from the module on personal and social well-being of ESS Round 3 and measured emotions, well-being, and engagement. All items were presented in German, but came from three different ESS questionnaire versions (i.e., the Austrian, German, and Swiss questionnaire). Findings show that not every face value difference in translations is critical in terms of measurement, but also that problems exist with emotional and effective words that can be translated in different ways with different statistical outcomes. However, split-ballot experiments exclusively conducted on the basis of translated texts do not necessarily solve the issue of comparability to the source text, which needs to be tackled in a next step. The author concludes that more in-depth analysis is underway and issues a cautionary note saying that different results on the item level do not automatically mean that the scale as a whole does not work.

References

ACL. (2017). The Association for Computational Linguistics. Retrieved from https://www.aclweb.org/portal/#page=39

Aitchison, J., Gilchrist, A., & Bawden, D. (2000). *Thesaurus construction and use: a practical manual*. Psychology Press.

Alcina, A. (2008). Translation technologies: Scope, tools and resources. *Target. International Journal of Translation Studies*, *20*(1), 79–102.

Anaphraseus. (2017). Anaphraseus. Retrieved from http://anaphraseus.sourceforge.net/

Badia, A. (2017). Technology in translation studies and translation assessment. Barcelona, Spain: First SERISS Symposium on synergies between survey translation and developments in translation sciences.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. In *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)*.

Baisa, V., Michelfeit, J., Medved, M., & Jakubícek, M. (2016). European Union Language Resources in Sketch Engine. In *LREC*.

Balkan, L., Jääskeläinen, T., Frentzou, C., & Kappi, C. (2011). European Language Social Science Thesaurus (ELSST): Issues in designing a multilingual tool for social science researchers. In *NEALT Proceedings Series*. Vol. 12.

Behr, D. (2017). Technological and computational advances in translation research and industry. Barcelona, Spain: First SERISS Symposium on synergies between survey translation and developments in translation sciences.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*(Feb), 1137–1155.

Berberich, K., & Kleiber, I. (2017). Tools for Corpus Linguistics. Retrieved from https://corpus-analysis.com/

Boitet, C. (1997). Machine-aided human translation. In G. B. Varile & A. Zampolli (Eds.), *Survey of the state of the art in human language technology* (pp. 257–261). Cambridge University Press.

Bojar, O., Buck, C., Federmann, C., Haddow, B., Koehn, P., Leveling, J., … Saint-Amand, H. (2014). Findings of the 2014 workshop on statistical machine translation. In *Proceedings of the ninth workshop on statistical machine translation* (pp. 12–58).

Bowker, L., & Fisher, D. (2010). Computer-aided translation. *Handbook of Translation Studies*, *1*, 60–65.

Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: {A} Practical Guide to Using Corpora*. Routledge.

Brown, P. F., Pietra, V. J. Della, Pietra, S. A. Della, & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, *19*(2), 263–311.

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

Burchardt, A., Richard, A. L., Rehm, G., Sasaki, F., Van Genabith, J., & Uszkoreit, H. (2014). Language Technology Drives Quality Translation. *MultiLingual*, *143*(4), 33–39.

Carson-Berndsen, J., Somers, H., Vogel, C., & Way, A. (2009). Integrated language technology as part of next generation localisation. *Localisation Focus*, *8*(1), 53–66.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *EMNLP 2014*. Retrieved from http://arxiv.org/abs/1406.1078

Christensen, T. P., & Schjoldager, A. (2016). Computer-aided translation tools–the uptake and use by Danish translation service providers. *Journal of Specialised Translation*, *25*, 89–105.

Clark, A., Fox, C., & Lappin, S. (2013). *The handbook of computational linguistics and natural language processing*. John Wiley & Sons.

Claverie, A. (2014). Post-editing MT: Is it worth a discount. *Multilingual*, *25*(3), 40–41.
Costa-Jussà, M. R., & Farrús, M. (2014). Statistical Machine Translation Enhancements Through Linguistic Levels: A Survey. *ACM Comput. Surv.*, *46*(3), 42:1--42:28. https://doi.org/10.1145/2518130
Costa-Jussa, M. R., & Fonollosa, J. A. R. (2015). Latest trends in hybrid machine translation and its applications. *Computer Speech & Language*, *32*(1), 3–10.
Costa-Jussa, M. R., Rapp, R., Lambert, P., Eberle, K., Banchs, R. E., & Babych, B. (2016). *Hybrid Approaches to Machine Translation*. Springer.
De Mulder, W., Bethard, S., & Moens, M.-F. (2015). A survey on the application of recurrent neural networks to statistical language modeling. *Computer Speech & Language*, *30*(1), 61–98. https://doi.org/https://doi.org/10.1016/j.csl.2014.09.005
Deepl. (2017). Deepl Translator. Retrieved from https://www.deepl.com/
Directorate-General for Translation. (2017). eTranslation. Retrieved from https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation
Eisele, A., Federmann, C., Saint-Amand, H., Jellinghaus, M., Herrmann, T., & Chen, Y. (2008). Using Moses to integrate multiple rule-based machine translation engines into a hybrid system. In *Proceedings of the Third Workshop on Statistical Machine Translation* (pp. 179–182). Association for Computational Linguistics.
España Bonet, C., Màrquez Villodre, L., Labaka, G., Díaz de Ilarraza Sánchez, A., & Sarasola Gabiola, K. (2011). Hybrid machine translation guided by a rule-based system. In *Machine translation summit XIII: proceedings of the 13th machine translation summit, September 19-23, 2011, Xiamen, China* (pp. 554–561).
European Social Survey. (2016). *ESS Round 8 Translation Guidelines*. London: ESS ERIC Headquarters.
Fantinuoli, C., & Zanettin, F. (2015). Creating and using multilingual corpora in translation studies. In C. Fantinuoli & F. Zanettin (Eds.), *New directions in corpus-based translation studies* (pp. 1–10). Berlin, Germany: Language Science Press Berlin.
Fantinuoli, C., Zanettin, F., Doms, S., Fotopoulou, A., Lapshinova-Koltunski, E., Mouka, E., … Sanz, Z. (2015). *New directions in corpus-based translation studies*. Language Science Press.
Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 303–315.
Folaron, D. (2010). Translation tools. *Handbook of Translation Studies*, *1*, 429–436.
Forcada, M. L., Tyers, F. M., & Ramírez Sánchez, G. (2009). The Apertium machine translation platform: five years on. *Proceedings of the 1st International Workshop on Free/Open-Source Rule-Based Machine Translation*, 3–10.
Forsyth, B. H., Kudela, M. S., Levin, K., Lawrence, D., & Willis, G. B. (2007). Methods for Translating an English-Language Survey Questionnaire on Tobacco Use into Mandarin, Cantonese, Korean, and Vietnamese. *Field Methods*, *19*(3), 264–283. https://doi.org/10.1177/1525822X07302105
Garcia, I. (2007). Power shifts in web-based translation memory. *Machine Translation*, *21*(1), 55–68.
Garcia, I. (2012). Machines, translations and memories: language transfer in the web browser. *Perspectives*, *20*(4), 451–461.
Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken: Wiley & Sons.
Harkness, J. A. (2005). SHARE translation procedures and translation assessment. In A. Borsch-Supan & H. Jurges (Eds.), *The Survey of Health, Aging, and Retirement in Europe - Methodology* (pp. 24–27). Mannheim, Germany: Mannheim Research Institute for the Economics of Aging (MEA).
Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, Adaptation, and Design. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 115–140). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470609927.ch7
Hutchins, W. J., & Somers, H. L. (1992). *An introduction to machine translation* (Vol. 362). London: Academic Press.

Kalchbrenner, N., & Blunsom, P. (2013). Recurrent Continuous Translation Models. In *EMNLP* (Vol. 3, p. 413).
Kietzmann, D., Wiehn, S., Kehl, D., Knuth, D., & Schmidt, S. (2016). Migration background and overall satisfaction with pre-hospital emergency care. *Applied Nursing Research*, *29*, 96–100. https://doi.org/10.1016/j.apnr.2015.05.009
Klein, G., Kim, Y., Deng, Y., Senellart, J., & Rush, A. M. (2017). OpenNMT: Open-Source Toolkit for Neural Machine Translation. *arXiv Preprint arXiv:1701.02810*.
Koby, G. S. (2013). Post-Editing of Machine Translation. *The Encyclopedia of Applied Linguistics*.
Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *MT summit* (Vol. 5, pp. 79–86).
Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., … Herbst, E. (2007). Moses: {O}pen source toolkit for statistical machine translation. In *Proceedings of ACL-2007* (pp. 177–180).
Lagoudaki, E. (2006). Translation memories survey 2006: Users' perceptions around TM use. In *proceedings of the ASLIB International Conference Translating & the Computer* (Vol. 28, pp. 1–29).
LES. (2017). Language Effects in Surveys. Retrieved from https://les.gu.se/
Linguee. (2017). Linguee. Retrieved from https://www.linguee.com/
Martens, M. (2017). *Uploaded and modularized TMT* (SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221 No. 3.12).
Merkel, M. (1993). When and Why Should Translations be Reused. In *Theory of Translation and Computers* (pp. 139–149). Vaan: 18th UAAKI Symposium on LSP.
Microsoft Corporation. (2017). Microsoft Translator.
Mohler, P. P., Dorer, B., De Jong, J., & Hu, M. (2016). Translation. In Survey Research Center (Ed.), *Guidelines for Best Practice in Cross-Cultural Surveys* (Second, pp. 233–377). Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from http://ccsg.isr.umich.edu/index.php/chapters/translation-chapter
Murray, G. C., Dorr, B. J., Lin, J., Hajič, J., & Pecina, P. (2006). Leveraging reusability. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL - ACL '06* (pp. 945–952). Morristown, NJ, USA: Association for Computational Linguistics. https://doi.org/10.3115/1220175.1220294
Nielsen, S. (2010). Specialized translation dictionaries for learners. In P. A. Fuertes-Olivera (Ed.), *Specialised Dictionaries for Learners* (pp. 69–82). Lexicographica.
O'Hagan, M. (2009). Computer-aided translation (CAT). In M. Baker & G. Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies* (pp. 48–51). London.
OECD. (2014). Translation and Adaptation Standards. In *PIAAC Technical Standards and Guidelines* (p. 201). Retrieved from http://www.oecd.org/skills/piaac/PIAAC-NPM(2014_06)PIAAC_Technical_Standards_and_Guidelines.pdf
OmegaT. (2017). OmegaT - The Free Translation Memory Tool - OmegaT. Retrieved from https://omegat.org/
OpenTM2 Steering Committee. (2017). OpenTM2 open source program for managing and handling global translation projects. Retrieved from http://www.opentm2.org/
Oxford University Press. (2017). Oxford English Corpus.
ProZ.com. (2011). *State of the industry: freelance translators in 2011*. Retrieved from http://www.proz.com/industry-report
Przepiórkowska, D., & Behr, D. (2017). *European Values Study Translation Guidelines*. Mannheim, Germany.
Pym, A. (2011). What technology does to translating. *The International Journal for Translation and Interpreting Research*, *3*(1).
Rinker, T. W. (2014). {regexr}: Readable Regular Expressions. Buffalo, New York. https://doi.org/10.5281/zenodo.13496
Schubert, L. (2015). Computational Linguistics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring, 20). Metaphysics Research Lab, Stanford

University.
SDL PLC. (2017). SDL Trados. Retrieved from http://www.sdltrados.com/
Serbina, T., Niemietz, P., & Neumann, S. (2015). Development of a keystroke logged translation corpus. In C. Fantinuoli & F. Zanettin (Eds.), *New directions in corpus-based translation studies* (Vol. 1, pp. 11–33). Berlin: Language Science Press.
Sha, M., & Lai, J. (2016). A case study of improving and evaluating consumer survey translation. *Translation & Interpreting*, *8*(1), 86–100.
Sin-wai, C. (2017). *The future of translation technology: towards a world without Babel* (First). Routledge New York. https://doi.org/10.1007/s10590-017-9199-x
Slocum, J. (1986). Computational Linguistics. In *American Journal of Computational Linguistics* (Spring 201, Vol. 12). Edward N. Zalta.
Somers, H. (2003). *Computers and translation: a translator's guide* (Vol. 35). John Benjamins Publishing.
Souto, M., Lupsa, D., & Dept, S. (2017). Survey translation and assessment possibilities for incorporating automated procedures. Barcelona, Spain: First SERISS Symposium on synergies between survey translation and developments in translation sciences.
Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufis, D., & Varga, D. (2006). The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. *arXiv Preprint cs/0609058*.
Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104–3112).
SYSTRAN. (2017). Systran. Retrieved from http://www.systransoft.com/systran/
Taravella, A., & Villeneuve, A. O. (2013). Acknowledging the needs of computer-assisted translation tools users: the human perspective in human-machine translation. *The Journal of Specialised Translation*, *19*(January), 62–74.
Varile, G. B., & Zampolli, A. (1997). *Survey of the state of the art in human language technology* (Vol. 13). Cambridge University Press.
Zavala-Rojas, D., Saris, W. E., & Gallhofer, I. N. (2018). Preventing Differences in Translated Survey Items using the Survey Quality Predictor. In T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC)*. New York: Wiley Series in Survey Methodology.