Deliverable Number: 3.8

Deliverable Title: Computational and corpus linguistic tools in questionnaire translation: guidelines to adopt and test for innovations in survey translation

Work Package: 3. “Maximising Equivalence through Translation”

Deliverable type: Report

Dissemination status: Public

Submitted by: Universitat Pompeu Fabra, European Social Survey ERIC

Author: Diana Zavala-Rojas

Reviewed by: Brita Dorer, GESIS-Leibniz Institut für Sozialwissenschaften

Date Submitted: May 2019

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, D (2019) *Computational and corpus linguistic tools in questionnaire translation: guidelines to adopt and test for innovations in survey translation*. Deliverable 3.8 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

Computational and corpus linguistic tools in questionnaire translation: guidelines to adopt and test for innovations in survey translation

Diana Zavala-Rojas
European Social Survey ERIC

Summary

The analysis of natural language by computational means has evolved rapidly in the last two decades. Survey practitioners and researchers can benefit from some applications that computational linguists have developed, especially those in the area of Computer-Assisted Translation (CAT).

In SERISS Task 3.2 we investigated and reviewed current developments in natural language processing, computational linguistics and corpus linguistics that have applications in the translation industry. Our aim was to find tools or methods that could eventually improve survey translation practice. As a result of this research, we suggest focusing on two areas of interplay between survey translation i.e. the translation and translation assessment of survey questionnaires, and the aforementioned scientific disciplines. Those areas are: 1) preparing survey items' texts for Computational Linguistics (CL) methods and translation technologies by the creation of multilingual parallel corpora and 2) enhancing procedures in survey translation with functionalities commonly present in commercial translation technologies. Academic survey practice is carried out through validated methodological procedures, innovations in the survey lifecycle should be tested. Therefore, in this report, we describe briefly the first area, creation of parallel corpora, but focus on providing guidelines on how to introduce innovations in translation processes.

A key aspect of survey research is to investigate how the 'survey method' plays a role in the respondent affecting their answers to a survey item. This area of survey research has been called *Cognitive Aspects of Survey Methodology* (CASM) (Schwarz, 2007). A typical example of this research is a split-ballot experiment testing among response scale formulations (Zavala-Rojas, Tormos, Weber, & Revilla, 2018). In this report we show how to test new methods and tools in the specifics of survey translation following the principles of CASM research, that is, by the means of (quasi)experimental events designed to test the impact of such new methods/tools.

We suggest implementing a three-stage methodology to systematically assess how the adoption of translation technologies and new methods would affect the quality of translations and translators' workflow. With this methodology we aim to provide guidelines for survey practitioners to design and plan (quasi)experiments to test innovations in their translation procedures, our guidelines follow good-practice in experimental research.

We provide a case study illustrating how to implement this methodology. The example describes important considerations to plan and design a pilot study to test computer-assisted translation software, machine translation and postediting activities in the translation process.

1. Introduction

The analysis and processing of (natural) language by computational means has evolved rapidly in the last two decades. When survey questionnaires are translated, survey research teams can benefit from some applications that computational linguists have developed, especially those included in computer-assisted or computer aided translation (CAT) tools –or *translation technologies* (terms that we will use interchangeably). Computational linguistic research and translation studies are very broad in scope. In Section 2, we contextualise some areas of language research for a survey research audience. We describe two areas of potential synergies between survey translation research and computational analysis of language. These areas can impact positively questionnaire translation practice: 1) Creating text data from translated survey questionnaires i.e. parallel corpora, to facilitate the implementation of linguistic methods and utilisation of CAT tools in the translation process, and 2) 'Enhancing current procedures in survey translation' with the adoption of translation tools commonly utilised in the translation industry. We define these areas and their key terms.

In addition to contextualise linguistic research for survey practitioners, the second main objective of this report is to present a *roadmap* to incorporate some methods and tools developed by computational linguists in *survey translation*, i.e. the translation and translation assessment of survey questionnaires. A roadmap is understood in the sense of guidelines. We suggest testing for the effect of introducing new tools in the translation process using (quasi)experiments[1]. To design such (quasi)experiments, we present a three-step methodology. Following those steps will simplify the design of empirical studies assessing translation process change. We use a case study to exemplify how to implement the three-step methodology in practice in a large-scale survey project

[1] An experiment is a scientific procedure carried out to test a hypothesis. Experiments allow to analyse cause-and-effect mechanisms by manipulating factors (treatment) from a clearly defined baseline situations (control) and observing the outcomes. An experiment is a fully controlled event designed for the purposes of hypothesis testing. We use the term (quasi)experiment to the procedures we describe in this report. Quasi experiments, in general, fail to control some features, for instance, random selection and assignment of participants to the treatment and control groups. (Quasi)experiment is a more appropriate term for experimental testing in the social sciences, as fully controlled, laboratory events are difficult to implement. See for reference (Coleman, 2018; Lawrence Neuman, 2011)

Section 3, Section 4 and Section 5 focus on the description and implementation of the three-step methodology to test translation process change using (quasi)experiments. We aim to provide guidelines for survey practitioners and researchers to plan and implement (quasi)experimental research to test innovations in their translation procedures. In order to achieve this goal, we use a case study as an illustration. We describe important considerations to design a pilot study to test the effect of CAT tools, 'Machine Translation' and 'postediting' activities in a large scale survey project. In our example, we assume that the translation process is based on the Translation, Review, Adjudication, Pretesting and Documentation (TRAPD) approach to survey translation (Harkness, 2003), as it is considered best-practice in survey translation. This case study can be taken as a footprint to conduct research about translation process change. Finally, in section 6 we provide concluding remarks.

2. Language research in context: which tools and methods can be incorporated in survey translation?

In the context of language research and translation practice, the acronym CAT refers to language translation in which a person, a *human translator*, utilises *application software* i.e. end-user computer programs, specifically designed to support translation tasks and manage translation processes (Garcia, 2007, 2012; O'Hagan, 2009; Pym, 2011; Taravella & Villeneuve, 2013). CAT tools are commonly used in most modern-day professional translation, making human translation without the aid of specialised software rather uncommon (Sin-wai, 2017). The development of CAT tools is highly interrelated with *computational linguistics* (CL), a broad interdisciplinary research field which applies statistical modelling and computational methods to natural language processing (NLP) (ACL, 2017). Corpus linguistics is the study of natural language expressed in texts, and the interrelation of texts in different languages. Often, corpus and computational linguists explore how to create text data, and how automatize gathering text data. Corpus methodologies allow empirical (statistical) approaches to translation studies (Fantinuoli et al., 2015; Oakes, Defeng, Ji, & Hareide, 2016).

Recently, a large variety of computer programs and methods to conduct tasks aimed at processing and analysing human languages have been developed, and many applications are transformed into functionalities included in software used by *language services providers* (LSP) i.e. translation researchers, translators, translation agencies. As a result, CL and corpus linguistics have changed the translation practice landscape. Adopting software to conduct translation activities implied that the translation industry transitioned from human-based into technology-oriented procedures.

Research in CL, corpus linguistics and translation studies is very broad in scope; lessons learned in the translation industry can be useful for survey practitioners and researchers to update survey translation practice. In the translation industry, introducing new methods and technologies has brought radical changes in procedures, leading to the adaptation of the workflow of translation teams to meet quality standards of translation projects. For instance, introducing the source texts in translation technologies was soon seen as a key step before aiding translators' work with technology. At present, CAT tools process several file formats, but this was not the case in the first versions of many technologies. Source texts arrived at LSP in a variety of formats, and an important part of CL methods dealt with processing texts to transform them into plain texts, i.e. text without format. Unformatted texts are needed in algorithms and software with translation functionalities.

Therefore, in this research project we focused on two areas with potential of directly connecting survey translation and computational analysis of language. Those areas are: 1) Creating text data from translated survey questionnaires i.e. parallel corpora, to facilitate the implementation of linguistic methods and the utilisation of CAT tools in the translation process, and 2) 'Enhancing current procedures in survey translation' with the adoption of translation tools commonly utilised in the translation industry. In the next subsections, we define those two areas and their key concepts (for an comprehensive review of linguistic tools and methods see Zavala-Rojas, 2017).

2.1 Multilingual parallel corpora of survey questions

Methods in CL, corpus linguistics and common applications in translation technologies work based on text corpus as input. Survey questions should be transformed into domain-specific corpora to study them from the perspective of CL. A (text) corpus is a large and structured database of texts, in which linguistic units are frequently parsed and annotated with metadata such as linguistic (e.g. grammar) and extra-linguistic characteristics (e.g. author). A corpus is the basic dataset to conduct linguistic quantitative analysis and to process natural language by computational means (Fantinuoli et al., 2015; Oakes et al., 2016).

A corpus is made up of texts in machine readable format, a Portable Document Format (PDF) file, or formatted text from a Microsoft Word file are not machine readable, they should be pre-processed to become text data. Source and target questionnaires in survey research are often formatted in a word editor, e.g. Microsoft Word, or transformed into a PDF file. These commonly used text formats need to be transformed into plain texts to create a corpus. An area of extensive research is *pre-editing* of texts, that is, the preparatory work of alignment, annotation and formatting of the source and target texts to convert them to text data (Serbina, Niemietz, & Neumann, 2015).

Once a corpus is created, texts have become data and it is possible to study some linguistic aspect from a quantitative perspective, for instance, the occurrence of words or the diversity of the vocabulary in the database. A common data form in computational text analysis is the so-called *document-term matrix (**X**)*. It is a bidimensional matrix that describes terms' frequency, usually words or pairs of words, in a *document*. Formally, a document-term matrix is defined as,

$\boldsymbol{X} = N \times P$, *where*

N is the number of documents, P the number of terms such that $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iP})$

A *document* in this context is an entry or record in the corpus, it can be any text of interest depending on the aim of the analysis, for instance a story or a press release, in the analysis of media data; a survey item, in the context of survey research, or a publication in Twitter or Facebook, in the context of digital social networks research. A *document-term matrix* allows, among others, using algorithms for classification of texts into topics, sentiment analysis e.g. how negative or positive a text is, or ranking texts based on search patterns of specific word occurrences.

Bi- or multi-lingual parallel corpora refer to datasets of texts that are aligned and, frequently, annotated (e.g. author, date, language, document version). Texts are subdivided into compositional and sequential units (terms) which can be retrieved as pairs (or triplets, etcetera) (Fantinuoli et al., 2015). Examples include the Europarl parallel corpora (Koehn, 2005), the EUR-Lex Corpus (Baisa, Michelfeit, Medved, & Jakubícek, 2016), the Acquis Communautaire (Steinberger et al., 2006) and the Norwegian-Spanish Parallel Corpus (NSPC) (Hareide, 2013). Frequently, parallel corpora include human translations, such as the Linguee database (Linguee, 2017), but it is also possible to create multilingual corpora that are not translations, but monolingual texts on a same subject in different languages, this is the case for the Spoken Wikipedia Corpora created with voice files of articles from the English, German, and Dutch Wikipedia that were recorded (Baumann, Köhn, & Hennig, 2018). Creating parallel corpora can be considered a problem of data engineering because the subdivision of texts in units conditions how language pairs can be retrieved. For example, a useful application of corpora of survey items would be to retrieve units that are translations of one specific response scale in several languages.

2.2 Enhancing survey translation with translation technologies

Although to different degrees, large scale survey projects have started to make use of translation technologies. Some examples of specialised translation software are the Open Language Tool (OLT) utilised in the translation process of the Programme for International Student Assessment (PISA) (OECD, 2012; PISA, 2010); the WebTrans software used in the translation process of the European Values Study 2008 (European Values Study & GESIS Data Archive for the Social Sciences, 2010) and the Translation and Management Tool (TMT) used by the Survey of Health, Ageing and Retirement in Europe (SHARE), at present in a pilot phase in the European Social Survey (ESS) and the European Values Study (EVS) (Martens, 2017). Except for the TMT, there is little publicly available information on the performance of specialised translation software in survey projects.

In the translation industry, specialised software has influenced enormously translators' work, as they must be familiar with different software functionalities for a variety of tasks. In the remaining of this section we define some of the most common features in CAT tools.

2.2.1 Translation memories

Cross-lingual text retrieval methods were one of the first consolidated CL applications in CAT technologies. A translation memory (TM) is a specific type of parallel corpus that facilitates the reuse of already translated segments. A TM can be considered a user-interface implementation of a corpus, with functionalities to extract, store, edit and/or reuse texts and/or translations, whereas a corpus is a structured database of such texts (translations) for a variety of purposes. A TM is a database made up of (validated) translated segments and aligned source and target texts that translators can access for re-using previous translations. Thus, while translating, the translator is provided with translation proposals, the so-called '*matches*', which they can annotate, accept, reject or revise (Sin-wai, 2017). TM systems can retrieve 100% matches, segments of text that match entries in the database exactly, or use fuzzy matching algorithms to retrieve similar segments, which are typically presented to the translator by flagging differences and indicating the matching percentage. A TM is usually accessed interactively from a CAT tool. The effectiveness of a TM matching algorithm can be enhanced via rule-based writing of text. Adhering to criteria for formatting and layout is also useful to have a correct segmentation of texts in user's interface. TM are still the most frequently functionality of CAT tools in professional translation (e.g. Lagoudaki, 2006; ProZ.com, 2011), alongside with *translator editors, terms' databases* and *text aligners*. *Machine translation and postediting* features have recently become common. Postediting refers to the editions that human translators implement to the machine translation outputs to obtain the final translations.

2.2.2 Translation editors

A basic component of CAT tools is a translation editor, which can interact with methods for cross-lingual text retrieval, and term databases. Common functionalities in a translator editor are those of word processing software e.g. Microsoft Word, including spell checking or copy/paste. A common task in a translation editor is to show a pre-translation of the text of interest based on a TM or machine translation tool, and to allow annotations to include metadata in the corpus metadata (Carson-Berndsen, Somers, Vogel, & Way, 2009; Nielsen, 2010; Somers, 2003).

2.2.3 Term databases

Another component commonly integrated into a CAT technology is a database for terms search, also referred to as a translation glossary, lexicon or terminology collection. A term database consists of words (terms) and lexical items related to such terms e.g. synonyms, definitions and examples (Nielsen, 2010). Common implementations of term databases are controlled vocabulary tools, such as thesauri:

they are lists of words or 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 Languages Social Science Thesaurus (ELSST) (Balkan, Jääskeläinen, Frentzou, & Kappi, 2011), a multilingual thesaurus developed by the Consortium of European Social Science Data Archives (CESSDA). In the SERISS Task 3.3 the ELSST is utilised for indexing social science surveys (Balkan, 2018).

2.2.4 Text aligners

Text aligners are tools that segment texts and pair them with respect to some linguistic properties. For instance, alignment is commonly done at the sentence level in two languages. One-to-one sentence correspondence of translated texts is not always possible. Due to the specificities of languages and translation style, sentences in the source text may have several transformation procedures in the target language version, such as being split, merged, deleted, among others, therefore automatic alignment is not a trivial computational task. Some CAT tools include alignment features, open-source and open-access aligners also exist for a variety of languages, for instance the LF Aligner (Andras, 2018) is a user-friendly application software based on the Hunalign engine (Varga et al., 2007), the Hunalign is regarded as one best open-source available multilingual aligners at present.

2.2.5 Other common technologies

CAT technologies are made up of four basic components: term databases, translation memories, text aligners and translation editors. In addition, there are other common features of specialised translation software aimed at making translators' work more efficient (Bowker & Fisher, 2010). For instance, some CAT tools have features for implementing sequences of 'automated checks' to compare source and target text with respect to characteristics, such as punctuation, length, numbers, dates (Souto, Lupsa, & Dept, 2017).

CAT technologies integrating features to manage translation projects, and track and store drafts are widely accepted by LSP, as they facilitate teamwork. Translation professionals are commonly linked by a CAT tool that manages the project's workflow, and in some cases even invoicing and accounting of professionals.

2.2.5 Machine translation and postediting

Machine translation is a highly multidisciplinary subfield of CL whereby human translators, engineers, computer scientists, mathematicians, statisticians, and linguists share research objectives. We distinguish between machine translation 'tools', that is, software that conducts instant machine translation e.g. Google Translate interface, and machine translation 'systems', that is, modelling approaches for machine translation , e.g. the algorithm producing translations shown in the Google interface. Machine translation outputs are considered *pre-translations* or draft translations because, first, preparatory work on the texts is needed, including checking the spelling, compiling dictionaries, and adjusting the text format, and second, they are considered pre-translations of the source text because they should be revised by a human translator (Sin-wai 2017).

Until recently, the LSP community's acceptance of machine translation was low; acceptance is becoming more widespread since the emergence of the neural networks paradigm (Bahdanau, Cho, & Bengio, 2014; Sutskever, Vinyals, & Le, 2014). Burchardt et al. (2014, pp. 34–35) argue that although machine translation tools are gaining acceptance in the translation industry, future successful tools will still require features of automated recognition and classification of machine translation outputs in at least three categories: 1) Output that can be used as is 2) machine translation outputs that can easily be post-edited to meet quality specifications, and 3) Output that should be discarded.

Currently, a large amount of state-of-the-art research in applied translation focuses on empirically comparing human translation approaches against machine translation and postediting tasks in translation projects. Measurement of translator's cognitive effort with respect to postediting activities is an area of extensive research (Vieira, 2014). Evidence suggests that postediting activities related to correcting lexicon and word order require most cognitive effort (Vieira, 2016). Translation research also focus on evaluating the usability of machine translation outputs across languages. Doherty and O'Brien (2014) concluded that the usability of such outputs is, in fact, dependent on the language pairs. They studied a source language, English, and four target languages: Spanish, French, German, and Japanese, the latter showing a considerably lower usability of machine translation outputs.

At present, research efforts also focus on comparing translators' acceptance of machine translation *vis a vis* TM outputs, as TM are the most common translation technology. Quasi experimental research is designed to compare evaluations of translations carried out without pre-translations and using TM outputs or machine translation outputs as pre-translations. In general, results show that translators' productivity and translations quality is higher when machine translation or TM outputs are used than when translations are carried out without them (Guerberof Arenas, 2014; Moorkens & Way, 2016). However, machine translation and TM outputs are as good as the data they are based on. If a stored translation is of poor quality, and it is made accessible, the match in the TM or the machine translation output will be of poor quality. Ideally, TM and machine translation systems would use only high-quality existing translations. However, this requires investment in quality and validation (Badia, 2017; Behr, 2017; Burchardt et al., 2014).

3. How to incorporate new methods and technologies in survey translation?

The utilisation of translation technologies in survey translation projects has not been systematically evaluated or publicly documented, for instance, in published survey guidelines or monographs. Although several survey projects have utilised specialised software in their translation processes there is little information about their effect in the project lifecycle and in the quality of the final translations.

Moreover, in recently published survey translation guidelines “any reduction of human involvement in the decision-making process of survey translation through an automatic mechanism” is discouraged because it can affect translation quality (Mohler et al. 2016), however no empirical findings are presented to support that claim. In the remaining of this section, we present a three-step methodology for testing translation process change by implementing quasi(experiments) in a survey project. Our aim is to provide user friendly guidelines that survey researchers can use to test innovations in their translation procedures.

The three steps that we suggest following to design a (quasi)experiment are:
Step 1. Identifying challenges in the translation process and translation tools to overcome such challenges
Step 2. Conducting a practical assessment of the translation process to be implemented in an (quasi)experiment
Step 3. Defining a pilot study

Experimental research should always be conducted under controlled conditions. That means that a research activity is organized specifically for the purposes of hypotheses testing. Step 1 and Step 2 allow to collect important information to specify a pilot study in detail taking important considerations into account.

This methodology aims at guiding researchers to design testing events of translation process change following good practices of experimental research. With these three-steps as a template, we aim to provide some basic guidelines to test for the impact of introducing innovations in translation procedures.

The method we outline follows the principles of *Cognitive Aspects of Survey Methodology* (CASM) (Schwarz, 2007) to design experiments investigating for method effects. Scholars in the area CASM research study *method effects*, by the means of testing events consisting of experiments. This allows to quantify how and to what extent the design of the questionnaire affects respondents, ultimately affecting data quality. A typical example of a research design in this area of statistical survey research is a split-ballot experiment testing the wording effects among several response scale formulations (e.g. Zavala-Rojas, Tormos, Weber, & Revilla, 2018). In the same way, translation experiments can be designed to test translation process change in a pilot study. We present a hypothetical case study to illustrate how the three-step methodology should be applied in practice in a survey project. The case study shows an example of an experimental design to test for the impact of introducing 'machine translation and postediting' activities in the translation process of a survey project. By using a case study we describe important considerations that researchers should take into account to design experiments about procedures in the lifecycle of a survey project.

3.1 Identifying challenges in the translation process

Survey researchers should aim at selecting CAT technologies/CL methods to improve current practice in their translation process, considering specific challenges. Therefore, the first step to design a pilot study is to identify challenges in the translation process. One and the same tool/method may be suggested for several problems.

Two key aspects should be clearly defined at this step:

1) Current challenges in the implementation of the questionnaire translation process.
2) Translation technologies or methods that have been helpful to overcome similar challenges in other areas of non-literary translation.

3.1.1 Case study: challenges of survey translation practice in large scale survey projects

Harkness (2003) suggested the so-called TRAPD method, a team/committee approach in a multistep process where different members provide their specific expertise to arrive at a final translation. Translation processes in large scale survey projects such as the European Social Survey (European Social Survey, 2016) or the Survey of Health, Ageing and Retirement in Europe (Harkness, 2005) are based on the TRAPD method[2]. It requires that at least two translators ('T' in the acronym) produce independent and parallel translations of the source version into a target language; or that source texts are split among two translators and thus all items are translated only by one translator. In a 'review' meeting, the translations are discussed by the two translators together with a 'reviewer' ('R'); an 'adjudicator' ('A') is responsible for the final decisions on

[2] Variants of the TRAPD method are used to translate questionnaires in cross-national surveys such as the European Social Survey (ESS) (European Social Survey, 2016), the European Values Study (EVS) (Przepiórkowska & Behr, 2017), the Survey of Health, Ageing and Retirement in Europe (SHARE) (Harkness, 2005), or the Programme for International Assessment of Adult Competencies (PIAAC) (OECD, 2014). Team approaches have also been applied for translation of questionnaires in market research (Sha & Lai, 2016), and in medical and health research (Forsyth, Kudela, Levin, Lawrence, & Willis, 2007; Kietzmann, Wiehn, Kehl, Knuth, & Schmidt, 2016).

different translation options[3]. The translated questionnaire is 'pretested' before fieldwork ('P') and the whole process is 'documented' ('D'). The team members should combine survey knowledge, translation and linguistic expertise, knowledge about the culture where the questionnaire will be administered, and potentially other knowledge related to the topic of the survey items.

We identify two potential challenges in the implementation of the TRAPD method in practice:
1) Familiarity with translation tasks of the translation team
2) Management and accessibility of user friendly documentation tracking translations versions throughout the process.

1) Familiarity with translation tasks of the translation team

While having a multi-disciplinary translation team is one of the TRAPD method advantages because team approaches provide several translation options to choose from, it can also be problematic when team members have different levels of training and experience in translating.

When survey researchers are part of the translation teams in the TRAPD method, they may not be very familiar with the methodology and the tasks. Usually, large-scale survey projects make internal guidelines available to national teams to aid this step, but translation teams may run out of time to exhaustively revise them. Academic high-quality social survey practice is highly demanding, the TRAPD may be carried out at the same time of other key fieldwork preparations in the survey lifecycle. Translation teams may lack resources to fully revise and implement guidelines for the translation process.

In the translation industry, in certain areas and for certain text types it has become common practice to carry out machine translation activities in combination with postediting activities/and human translation (Balling & Carl, 2014). 'machine translation and postediting' activities may potentially ease translators work by providing a high-quality pre-translation from which the translation team can generate a first draft (Sin-wai, 2017). Teams would have a starting point, potentially reducing the time and resources required at this step. Findings of test studies show that translators' productivity is enhanced, without compromising the final translation quality, when translators have a high-quality pre-translation to be post-edited (Guerberof Arenas, 2014).

2) End-user documentation tracking translations versions

[3] Harkness suggested that to reduce costs, the roles of reviewer and adjudicator can also be merged in one person (becoming a so-called 'reviewer-cum-adjudicator')

‘Documentation’ is an important dimension of the TRAPD method. The decision-making process to arrive to a final translation and commentary on translation versions should be documented. (Behr, Dept, & Krajčeva, 2018) categorised translation documentation as input documentation (texts that are prepared before translation, including source texts and guidelines) and output documentation (texts that are produced in the translation process, including draft and final translation versions). Managing, storing, analysing, and reusing this information in a systematic way is a potential challenge of the TRAPD method. A completed round of translation in a multilingual survey project would generate many input and output document files e.g. Excel spreadsheets. Although, it is possible to analyse translation documentation in a case-by-case basis (see for instance Behr et al., 2018; Mohler et al., 2010; Mohler & Uher, 2003), systematic analysis of survey translation documentation is not currently done to a large extent.

Integrated CAT tools are application software with features for managing translation projects. Some of them allow coordination of team approaches to translation and offer features for tracking translation versions and managing commentary of translation processes (see for instance: OmegaT, 2017; Ramírez Polo, 2013; Serbina, Niemietz, & Neumann, 2015).

3.2 Conducting a practical assessment

Conducting a *practical assessment* is the second stage in the three-step methodology to design a pilot study for translation process change. Firstly, research teams should identify concrete tools/methods to be tested. This step is not trivial, in Section 3 we summarised two broad areas in linguistic research and translation practice which can bring potential benefits to survey translation, but many concrete tools exist, for instance, the Linguist List (International Linguistics Community Online, 2018) contains about two hundred computer tools developed in the area of linguistics. The computer codes of a large amount of CL methods are publicly released as software modules that can be integrated, for instance, in open-source software. The NLP Stanford University repository (Manning, 2014), for example, includes a large list of downloadable computer programs (more than a hundred) for a variety of language processing and analysis tasks.

The second objective in a practical assessment is to identify in detail how specific translation activities would be modified or substituted by the incorporation of new methods and technologies. Anticipating to the maximum extent potential changes in the translation process allows specifying better the conditions of the experimental settings in the next stage. In order to compare performance, the present translation process should be recreated in the experiment. One group of participants should recreate the baseline situation and form the “control” group.

A third objective in a practical assessment is to decide the profile of the participants in the empirical study. Although the validity of an experiment is not dependent on the profile of the individuals taking part in the experiment, but on the definition of the experimental conditions (groups), their variations and the random assignment of cases, selecting the best available sample for the specific research question is important (see Lawrence Neuman, 2011 for a discussion about validity in experimental research).

For instance, in quantitative psychology, a discipline that makes an extensive use of (quasi)experiments for hypotheses testing, participants often are undergraduate or graduate students (Hanel & Vione, 2016). In survey research, participants in test studies are, in general, representative of the main population of interest. These individuals are part of the sampling frames. In translation studies, there are several possibilities. Some empirical studies have been conducted with trained translators; some others have used translators in training e.g. translation post-graduate students (see O'Brien, 2009 for a discussion on participants in translation research).

Participants in studies testing translation procedures in social surveys could be national translation teams already active in such procedures, professional translators with experience in translating questionnaires, or professional translators in training. All options have advantages and disadvantages. In general, considerations on what is the most important knowledge to be obtained in the test study should guide the decision on the profile of participants.

Finally, in a practical assessment step a *linguistic map* of the survey project should be summarised. Large scale survey projects are linguistically diverse, both in the number of languages spoken and in the regional differences. Therefore, selecting the languages of a pilot study is key to produce results that can be informative to survey practice. Although it would be ideal to test process change in all languages, it is very unlikely that a survey project has resources for a pilot study in every participating country or language version.

For instance, selecting languages to test 'machine translation and postediting' activities is not trivial. Doherty and O'Brien, (2014) have shown that usability of machine translation outputs is dependent on language, that is, they are not equally valuable across languages. Recent literature suggests that poor quality machine translation outputs can affect translators' cognitive effort (Doherty, 2015; Doherty & O'Brien, 2014; Gaspari, Almaghout, & Doherty, 2015) and final translations quality (Burchardt et al., 2014).

Additionally, discussions about translation options are, in general, carried out in the target language, thus although it does not affect the validity of the experiment, the linguistic proficiency of members of the research team i.e. the team coordinating, analysing and potentially publishing the results of the experiment, is an important pragmatic consideration to design a pilot study.

3.2.1 Case study: concrete CAT tools, machine translation and postediting tools.

For the case study presented in this report, two examples of open-source integrated CAT tools are Matecat (Federico et al., 2014) and OmegaT (OmegaT, 2017). Both have been previously used in research-oriented projects and allow for 'machine translation and postediting' functionalities. We strongly argue in favour of testing open-source CAT tools, since they offer several advantages, among others, that they can be customised to the specifics of survey translation projects, although this requires in-house programming skills in the survey project team. Another advantage is that the cost of software license and maintenance is considerably lower than in commercial tools.

Commercial online instant machine translation tools such as Deepl, Google Translate or Microsoft Translate allow translating texts, but they are not translation technologies optimized for postediting activities. If these tools were tested in a pilot study, postediting activities and documentation would require other software such as word editors, spreadsheets or tailored survey translation software, e.g. the TMT.

3.2.2 Case study: definition of additional activities introduced by new tools or methods

Having identified concrete tools, the next step to design a testing event using (quasi)experiments is to describe what components will built out the experiment by defining 'control' and 'treatment' activities.

A first possibility is to test a CAT software for translation and postediting activities in the translation process. An advantage of this scenario is that these tools are designed specifically for translation activities, a disadvantage is that the resulting translation process will require adapting several new activities. Two additional tasks are introduced: pre-editing the source texts, and postediting machine translation outputs.

A second possibility is to use an instant machine translation tool in combination with tools used at present in the survey project, for instance, a spreadsheet or an application software like the TMT. The advantage of this configuration is that the activities in the experiment would be similar to the present situation, only few changes would be introduced in the translation process. A disadvantage of testing 'machine translation and postediting' without CAT tools is that the translation process would still require handling several software to complete all activities.

Three additional tasks should be incorporated in the translation workflow in this scenario: pre-editing the source texts, that is, the task of introducing them in a machine translation tool, extracting those outputs from the tool and introducing them in the tool in which postediting activities will be conducted. Once the machine translation outputs are obtained, 'postediting' activities are all editions that human translators implement to obtain the final translations.

3.2.3 Case study: participants in the pilot study

In addition to the basic new activities, translation teams participating in the pilot study should be familiar not only with the translation guidelines of the survey project conducting the experiment, but also with the software that will be used in the pilot study. It is reasonable to expect that depending on the profile of the participants in the study, they will have a larger learning curve to get familiar with translation activities if they are survey experts, or if they are trained translators, their learning curve will be larger with respect to the specifics of survey questionnaires.

A research design with participants already involved in survey translation teams would be highly informative for survey practitioners, however, with this design it is not possible to disentangle to what extent lack of familiarity with translation tasks and technologies in the team members will affect results. A research design with only trained translators or -translators in training- will provide an estimate of the relative effect of introducing a new method or tool when the profile of the participants is constant. The disadvantage is that the conditions that originate the results may not be very similar to the specifics of survey practice.

3.2.4 Case study: linguistic diversity in the survey project and possibilities for language selection for a pilot study

Large scale survey projects are highly linguistically diverse[4], a successful pilot study should provide information about the effect of the new procedures and tools in several languages. There are several possibilities for language selection to conduct a pilot study. A first option is selecting languages for which versions are produced in several countries, for instance, in the ESS several country versions are produced for Russian, German, French, Dutch/Flemish and Italian languages. A second option is selecting languages from different linguistic families: for instance, Romance, Slavic and Western Germanic languages. A third option is selecting languages based on their degree of similarity with the source language, for instance, languages close to the English source language in SHARE are Danish, Swedish and Norwegian. ESS and SHARE languages that are distant from English are Arabic, Hebrew, Hungarian, Finnish and Polish.

3.3 Defining the pilot study

The third stage to design an experimental study about changes in survey translation process consists of defining the experimental conditions. Experimental research should be conducted under controlled circumstances, a research activity should be organized

[4] The ESS Round took place in 23 countries and SHARE Wave 7 took place in 28 countries.

specifically for the purposes of hypotheses testing. The information collected at Step 1 (Section 3.1) and Step 2 (Section 3.2) should allow specifying a pilot study in detail taking important considerations into account. Table 1 summarises the information collected for the case study in this report. This information is the basis to specify the operational aspects of the experiments including: 1) concrete activities that participants will conduct at the treatment and control environments, 2) concrete hypotheses and, 3) indicators to analyse the experiment.

We present here an example of hypotheses that could be tested in an implementation of the case study for the case of testing a CAT tool, machine translation and postediting activities:

1. *Introducing a CAT tool,* machine translation *and postediting activities have a positive effect on the quality of the final translations, and a 'negative' effect in the workload of the translators (reduces workload)*.

A second general hypothesis for the case when a CAT tool is not used, only 'machine translation and postediting' would be:

2. *The effect of an instant* machine translation *tool is positive on the quality of the final translations, and negative in the workload of the translators (reduces workload)*.

From these statements the dependent and independent variables are derived 'quality of final translation outputs' and 'translators' workload', although they are not specifically operationalised as indicators yet. The dependent variables are not overall measures, they should be defined for the purposes of the experiment. The unit of analysis is a survey item translated from English into a target language. The operationalisation of the dependent and independent variables as indicators is not trivial, there are several possibilities, we will describe considerations to select indicators in section 3.3.1 below.

Table 1. Summary of information that should be gathered at the theoretical and practical assessment stages

	THEORETICAL ASSESSMENT	**PRACTICAL ASSESSMENT**
	Which are the challenges in the translation process?	How would the translation process change by integrating new procedures/tools?
CASE STUDY	*Familiarity with translation tasks of the translation team (see Section 3.1.1)* *Management and accessibility of user-friendly documentation tracking translation versions throughout the process (see Section 3.1.1)*	*Additional activities with CAT software: getting familiar with CAT software, introducing source texts into the tool, obtaining machine translation (MT) outputs and postediting activities.* *With instant MT tools: introducing source texts into the MT tool, obtaining MT outputs, introducing outputs in the tool to conduct postediting activities. (See Section 3.2.2)*
	Which translation technologies have been useful to overcome those challenges in other	Which specific translation technologies can be tested?

	areas of translation?	
CASE STUDY	*MT and postediting (see Section 3.1.1)* *CAT software (see Section 3.1.1)*	*CAT software: MateCat, OmegaT, Instant MT tools: Deepl, Google Translate, Microsoft Translator (See Section 3.2.1)*
		Which is the ideal profile of participants for a pilot study?
CASE STUDY		*Two configurations of translation teams allow testing is the effect of the tools is dependent on the profile of the translators: 1) the same composition of translation teams as recommended for any TRAPD / committee/team approach: Survey researchers and trained translators, all experienced in reviewing and translating questionnaires. (See Section 3.2.3)*
		Which are the languages of the survey project?
CASE STUDY		*Approximately 30 language versions. (See Section 3.2.4)*

Table 2 shows a comprehensive 4x2 experimental design in three languages, considering all the information derived from Table 1. This design is a between-subject factorial design as several independent factors are manipulated simultaneously to observe their effects on the dependent variables. In this example, we assume that the new tools and methods are tested in three languages translated from English. Consequently, there are three control groups, one per language (Language A, Language B and Language C), which should implement the current translation procedure for a number *N* of items. The *experimental variations* are four alternatives to conduct translation activities with two profiles of participants (groups), the same number of items *N* are translated in each group.

An experimental design can easily get very complicated, as Table 2 shows. The design exemplified here would require 24 groups, in practice, it would be very difficult to implement it in a survey project. We present the full design as an illustration. When conducting experiments, survey research teams should choose those scenarios that can provide information to improve their translation practice.

Table 2. Summary of the experimental groups in the case study

		CONTROL GROUP (CURRENT PROCEDURE)	**TOOLS FOR MACHINE TRANSLATION AND POSTEDITING**		
PARTICIPANTS PROFILE	Language	Human translation in the documentation tool e.g. Excel spreadsheet	Machine Translation (MT) tool A + documentation tool	MT tool A + TMT	CAT tool A with MT feature
ONE SURVEY	English to	Experimental	Experimental	Experimental	Experimental

RESEARCHER AND ONE TRAINED TRANSLATOR	language A	group 1	group 4	group 7	group 10
	English to language B	Experimental group 2	Experimental group 5	Experimental group 8	Experimental group 11
	English to language C	Experimental group 3	Experimental group 6	Experimental group 9	Experimental group 12
TWO TRAINED TRANSLATORS.	English to language A	Experimental group 13	Experimental group 16	Experimental group 19	Experimental group 22
	English to language B	Experimental group 14	Experimental group 17	Experimental group 20	Experimental group 23
	English to language C	Experimental group 15	Experimental group 18	Experimental group 21	Experimental group 24

The implementation of this illustration can be as follows: In Group 1 to Group 3 human translation is conducted in the first step of the translation process usually implemented e.g. the TRAPD method; and documented in the usual tool, for instance in an Excel spreadsheet. The profile of the participants in the control groups is mixed, a survey expert and a translator conduct the first step of the TRAPD method.

A pilot study that tests if the effect of introducing CAT tools, machine translation and postediting activities is dependent on the characteristics of the participants, would require experimental groups with different profiles of participants (sometimes this design is called *between-subject design*). In the example, participants in Group 4 to Group 12 have a similar profile as in the control groups. The translation team is interdisciplinary, it has a survey expert and a trained translator with experience in translating questionnaires. Participants in Group 13 to Group 24 are two trained translators instead of a mixed team.

The first variation or treatment is conducted in Group 4 to Group 6: source questions feed an instant machine translation tool, e.g. Deepl, the outputs are then introduced in a spreadsheet to conduct postediting, translations versioning and commentary are stored in that tool. The second variation is represented by Group 7 to Group 9 and is very similar to the first one, the only change is that instead of the spreadsheet, the TMT software is used for postediting, tracking translations versioning and storing commentary.

In the third treatment condition, Group 10 to Group 12, participants conduct translation activities with a CAT tool that integrates machine translation features, e.g. OmegaT or Matecat. In this condition, translation is conducted in the following way: source texts are introduced in the CAT tool, machine translation outputs are obtained, and post editing activities are conducted in the same software, which handles translations versioning and commentary. Group 13 to Group 24 repeat the same pattern of experimental variations, but the profile of the participants is different.

Random assignment is one of the most important components of an experimental design. Therefore, participants should be randomly assigned to groups. As participants learn in the process of translating, the order of survey items to be translated should also be random (Lawrence Neuman, 2011). As we see in the example, a factorial design would require that participants are not shared by different groups, otherwise, learning effects will confound the results of the experiment.

3.3.1 Specifying the analysis strategy

Measuring the effect of a new tool or method in a process requires quantifying properties of that process. In the research design we present here, properties about the quality of the translations and translators' workload in the control group should be compared against those same properties in the treatments for each language. A common indicator to study the effect of 'machine translation and postediting' in the quality of a translation used in cognitive translation studies literature is the number of *editing operations* required to arrive at a final translation, a second frequent indicator are *subjective ratings of translations by external evaluators*. When external evaluators are used, they should not know which group originated the translation. Frequent indicators of translators' workload are *pauses* (eye tracking), *editing time*, and *subjective ratings of postediting tasks' effort* (Vieira, 2016). The analysis strategy consists on comparing such metrics between control and treatment groups in each language.

Studying the overall effect of introducing innovations in the translation process can combine qualitative and quantitative strategies. Qualitative methods can include debriefing of the 'TRA' steps and analysis of the committee meeting discussions (Behr, 2009). Quantitative indicators can be related to comparisons of total time of the translation activities, subjective ratings of translators' satisfaction, number of editions to arrive to a final translation. Mixed-method approaches combine both, quantitative and qualitative analysis (Dorer, forthcoming: 2019).

4. Conclusions

SERISS Task 3.2 consisted of investigating the suitability of tools used in CL, corpus linguistics and the translation industry to support survey translation. As methods and tools in language research are abundant, we focused on applications that could provide useful support in the context of translating questionnaires.

Sin-wai (2017) estimates that over 90 per cent of translation projects are of non-literary translation, and only 10 per cent is about literary translation. In this categorisation, the translation of survey questionnaires is considered non-literary translation. Therefore, lessons learned in the translation industry can be useful for survey researchers to update procedures. New methods and technologies brought radical changes in the translation industry, adding new steps or leading to the adaptation of procedures in the workflow. We suggest that survey researchers should focus on two areas of computational analysis of language and corpus linguistics to update survey translation practice: 1) preparing survey items' texts for CL methods and CAT tools by the creation of multilingual parallel corpora and 2) adopting functionalities commonly present in commercial translation technologies in survey translation and testing for the effect of these innovations.

Survey translation can be enhanced by introducing translation technologies and computational methods of processing language data after assessing the challenges and potential benefits of changing consolidated survey translation processes. Changes in translation approaches can be studied by the means of (quasi)experimental designs. Assessing the impact of incorporating new methods and technologies may represent a large investment of economic and human resources, but survey researchers would have information on the impact of process change and can rely on the know-how developed in other areas of translation and in the expertise developed in survey methodology to design experimental research.

One of the conclusions of this research project is that translation technologies and CL methods offer a variety of potential solutions to survey translation needs. However, the introduction of technology within the translation process requires an analysis at different dimensions. We suggested a three-step methodology to design a pilot study assessing how new procedures may work in practice.

An assessment is required to determine the rationale behind introducing new methods because it conveys a modification of human procedures (Step 1). Test studies should include a detailed assessment of operational aspects (Step 2). The definition of a pilot study using experimental designs should be very specific about the purposes of such experiments, only in this way effects of a process can be quantified (step 3).

We use a case study to describe the use of this three-step methodology in practice: we show how to design a pilot study testing the impact on translation quality and translators' workload by allowing the use of CAT software, machine translation and postediting activities to produce translations. Experimental research can easily get complicated and analysis can be difficult if research questions and the experimental conditions and are not very specific. In the example, the experiment requires 24 groups. This complexity should be avoided by having clear options that are informative for the survey team implementing the pilot study.

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.

Andras, F. (2018). *LF Aligner*. Retrieved from https://sourceforge.net/projects/aligner/

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*. Retrieved from http://arxiv.org/abs/1409.0473

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

Balkan, L. (2018). *Report on application of indexing terms in the data lifecycle*. Retrieved from www.seriss.eu/resources/deliverables

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. *NEALT Proceedings Series*. Vol. 12.

Balling, L. W., & Carl, M. (2014). *Post-editing of machine translation: Processes and applications*. Cambridge Scholars Publishing.

Baumann, T., Köhn, A., & Hennig, F. (2018). The Spoken Wikipedia Corpus collection: Harvesting, alignment and an application to hyperlistening. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-017-9410-y

Behr, D. (2009). Translationswissenschaft und international vergleichende Umfrageforschung (Vol. 2).

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.

Behr, D., Dept, S., & Krajčeva, E. (2018). Documenting the Survey Translation and Monitoring Process. *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, 341–355.

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

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.

Coleman, R. (2018). *Designing experiments for the social sciences: how to plan, create, and execute research using experiments*. Retrieved from http://us.sagepub.com/hi/nam/designing-experiments-for-the-social-sciences/book255818#.XMGT2fXYlvI.mendeley

Doherty, S. (2015). The Impact of Translation Technologies on the Process and Product of Translation. *International Journal of Communication*, *9*(February). https://doi.org/10.1300/J009v26n03_04

Doherty, S., & O'Brien, S. (2014). Assessing the Usability of Raw Machine Translated Output: A User-Centered Study Using Eye Tracking. *International Journal of Human-Computer Interaction*, *30*(1). https://doi.org/10.1080/10447318.2013.802199

Dorer, B. (forthcoming: 2019): Advance Translation as a Means of Improving Source Questionnaire Translatability? Findings from a Think-Aloud Study for French and German. Berlin: Frank & Timme

Dorer, B. (2011). *Advance translation in the 5th round of the European Social Survey*

(ESS) (No. 2011–4.). Lausanne.
European Social Survey. (2016). *ESS Round 8 Translation Guidelines*. London: ESS ERIC Headquarters.
European Values Study, & GESIS Data Archive for the Social Sciences. (2010). *EVS 2008 method report: documentation of the first data release in June 2010*. Retrieved from https://nbn-resolving.org/urn:nbn:de:0168-ssoar-207205
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.
Federico, M., Bertoldi, N., Cettolo, M., Negri, M., Turchi, M., Trombetti, M., … Martines, A. (2014). The matecat tool. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, 129–132.
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.
Gaspari, F., Almaghout, H., & Doherty, S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, *23*(3), 333–358. https://doi.org/10.1080/0907676X.2014.979842
Guerberof Arenas, A. (2014). Correlations between productivity and quality when post-editing in a professional context. *Machine Translation*, *28*(3–4). https://doi.org/10.1007/s10590-014-9155-y
Hanel, P. H. P., & Vione, K. C. (2016). Do Student Samples Provide an Accurate Estimate of the General Public? *PloS One*, *11*(12), e0168354–e0168354. https://doi.org/10.1371/journal.pone.0168354
Hareide, L. (2013). *The Norwegian-Spanish Parallel Corpus*. Retrieved from http://hdl.handle.net/11509/73
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).
International Linguistics Community Online. (2018). Linguist List - Text and Computer Tools Index Page. Retrieved December 3, 2018, from LINGUIST website: https://linguistlist.org/tools/
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
Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. *MT Summit*, *5*, 79–86.
Lagoudaki, E. (2006). Translation memories survey 2006: Users' perceptions around TM use. *Proceedings of the ASLIB International Conference Translating & the Computer*, *28*(1), 1–29.
Lawrence Neuman, W. (2011). Experimental research. In *Social research methods: Qualitative and quantitative approaches* (Seventh ed). Pearson.
Linguee. (2017). Linguee. Retrieved from https://www.linguee.com/
Manning, C. (2014). Statistical NLP / corpus-based computational linguistics resources. Retrieved November 14, 2018, from https://nlp.stanford.edu/links/statnlp.html
Martens, M. (2017). *Uploaded and modularized TMT* (No. 3.12).

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). Retrieved from http://ccsg.isr.umich.edu/index.php/chapters/translation-chapter

Mohler, P. P., Hansen, S. E., Pennell, B.-E., Thomas, W., Wackerow, J., & Hubbard, F. (2010). A Survey Process Quality Perspective on Documentation. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 299–314). https://doi.org/10.1002/9780470609927.ch16

Mohler, P. P., & Uher, R. (2003). Documenting comparative surveys for secondary analysis. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (Vol. 325, p. 311). Wiley-Interscience.

Moorkens, J., & Way, A. (2016). Comparing Translator Acceptability of TM and SMT Outputs. *Baltic J. Modern Computing*, *4*(2).

Nielsen, S. (2010). Specialized translation dictionaries for learners. In P. A. Fuertes-Olivera (Ed.), *Specialised Dictionaries for Learners* (pp. 69–82). Lexicographica.

O'Brien, S. (2009). Eye tracking in translation process research: methodological challenges and solutions. *Methodology, Technology and Innovation in Translation Process Research*, *38*, 251–266.

O'Hagan, M. (2009). Computer-aided translation (CAT). In M. Baker & G. Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies* (pp. 48–51). London.

Oakes, M., Defeng, L., Ji, M., & Hareide, L. (2016). Introduction to Corpus Methodologies Explained: An empirical approach to translation studies. In L. Defeng, L. Hareide, M. Ji, & M. Oakes (Eds.), *Corpus Methodologies Explained* (pp. 11–14). Routledge.

OECD. (2012). *PISA 2012 Technical Report*. Retrieved from http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

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/

PISA. (2010). *Translation and Adaptation Guidelines For PISA 2012*. Retrieved from National Project Managers' Meeting website: https://www.oecd.org/pisa/pisaproducts/pisa2012translationmanualsandguidelines.htm

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).

Ramírez Polo, L. (2013). Managing the translation workflow with a Computer Assisted Translation Tool: SDL Trados 2011. *Language Value*, *5*(1), 161–174. Retrieved from http://www.languagevalue.uji.es/index.php/languagevalue

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, *21*(2), 277–287.

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). In *Routledge Studies in Translation Technology* (First). https://doi.org/10.1007/s10590-017-9199-x

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. *Advances in Neural Information Processing Systems*, 3104–3112.
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.
Varga, D., Halácsy, P., Kornai, A., Nagy, V., Németh, L., & Trón, V. (2007). Parallel corpora for medium density languages. *Amsterdam Studies In The Theory And History Of Linguistic Science Series 4*, *292*, 247. Amsterdam; Philadelphia; J. Benjamins Pub. Co; 1999.
Vieira, L. N. (2014). Indices of cognitive effort in machine translation post-editing. *Machine Translation*, *28*(3–4). https://doi.org/10.1007/s10590-014-9156-x
Vieira, L. N. (2016). How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data. *Machine Translation*, *30*(1–2). https://doi.org/10.1007/s10590-016-9188-5
Zavala-Rojas, D. (2017). *Report of the state of the art on tools in computational linguistics and computer-aided translation for enhancing survey translation*. Retrieved from https://seriss.eu/resources/deliverables/
Zavala-Rojas, D., Tormos, R., Weber, W., & Revilla, M. (2018). Designing response scales with multi-trait-multi-method experiments. *Mathematical Population Studies*, *25*(2), 66–81. https://doi.org/10.1080/08898480.2018.1439241