

Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782

Start Date of Project: 01/01/2019

Duration: 40 months

Deliverable 4.9 Guidelines on the use of Translation Memories in survey translation

Dissemination Level	PU
Due Date of Deliverable	31/07/2021 (M31)
Actual Submission Date	19/07/2021
Work Package	WP4 Innovations in Data Production
Task	Task 4.3 Applying Computer Assisted Translation in Social Surveys
Type	Report
Approval Status	Waiting for EC approval
Version	V2.2
Number of Pages	p.1 – p.17

Abstract:

This report introduces the concepts of computer-aided translation environments and Translation Memories to social scientists. Furthermore, it documents a possible use case scenario of Translation Memories for questionnaire translation. Guidelines on the use of Translation Memories are provided in the form of a recorded demo.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782

History

Version	Date	Reason	Revised by
0.0	29/04/2021	First draft sent to WP 4 GESIS team leader	Dorothee Behr
0.1	06/05/2021	WP 4 GESIS team leader revises content	Dorothee Behr
0.2	07/05/2021	WP 4 GESIS colleague revises content	Brita Dorer
0.3	10/05/2021	Author revises the draft	Veronika Keck
1.0	26/05/2021	Authors send formatted version to peer review	Dorothee Behr
1.1	15/06/2021	Peer reviewed	Patricia Goerman
1.2	15/06/2021	Peer reviewed	Anna Andreenkova
1.3	15/06/2021	Author addresses peer review comments	Veronika Keck
1.4	16/06/2021	WP 4 GESIS team leader revises content	Dorothee Behr
1.5	21/06/2021	WP 4 GESIS colleague reviews content	Brita Dorer
2.0	30/06/2021	Final version reviewed by WP4 leader (UPF)	Diana Zavala-Rojas
2.1	05/07/2021	Author revises the draft	Veronika Keck
2.2	09/07/2021	WP 4 GESIS team leader clarifies with author remaining issues	Dorothee Behr

Author List

Organisation	Name	Contact Information
GESIS	Veronika Keck	veronika.keck@gesis.org
GESIS	Dorothee Behr	dorothee.behr@gesis.org
GESIS	Brita Dorer	brita.dorer@gesis.org

Executive Summary

Task 4.3 in WP4 *Innovations in Data Production* of the SSHOC project is dedicated to *Applying Computer-Assisted Translation tools in Social Surveys*. A key activity of this task is to incorporate newly created Translation Memories (TMs) from a corpus, which has been developed in Task 4.2 (*Preparing tools for the use of Computer Assisted Translation*), into an open-source computer-assisted translation (CAT) environment. Moreover, this report lays out a test case to demonstrate the feasibility of the usage of TMs within a CAT environment. The overarching goal is to introduce CAT tools and their usage possibilities to survey researchers to increase the awareness and adoption of existing and new digital tools for the survey translation process.

Using the CAT tool MateCat, this report provides a theoretical framework of CAT tools with a focus on TMs, their practical use and a summary of the advantages and disadvantages of TM usage. To support social scientists in their increasingly digitized translation processes, increase awareness, and showcase the usability of TMs, a demo of the integration of TMs into the MateCat translation environment was recorded. The demo shows how the public TM MyMemory and private TMs created based on previously translated international surveys can be used and updated to increase the efficiency and quality of the survey translation process.

Abbreviations and Acronyms

CAT tool	Computer-assisted translation tool
ESS	European Social Survey
EVS	European Values Study
MCSQ	Multilingual Corpus of Survey Questionnaires
MT	Machine translation
TM	Translation Memory
SHARE	Survey of Health, Ageing and Retirement in Europe
SSHOC	Social Sciences and Humanities Open Cloud

Table of Contents

1.	<i>INTRODUCTION</i>	6
2.	<i>MATECAT AND TRANSLATION MEMORY.....</i>	7
3.	<i>PRODUCING TMX FILES FROM THE MULTILINGUAL CORPUS OF SURVEY QUESTIONNAIRES.....</i>	9
4.	<i>INTEGRATION OF TRANSLATION MEMORIES INTO A MATECAT PROJECT AND THEIR USE</i>	10
5.	<i>ADVANTAGES AND DISADVANTAGES OF TRANSLATION MEMORIES IN GENERAL</i>	11
6.	<i>ADVANTAGES AND DISADVANTAGES OF TRANSLATION MEMORIES FROM THE MULTILINGUAL CORPUS OF SURVEY QUESTIONNAIRES.....</i>	12
7.	<i>CONCLUSION</i>	14
8.	<i>REFERENCES</i>	16

1. Introduction

Various disciplines conduct surveys to obtain research data. Against the backdrop of globalization and interconnectivity, survey research across these disciplines has increasingly become cross-national and cross-cultural. Many factors influence the comparability of data from cross-national or cross-cultural survey research. One of them is high-quality translation. To ensure high-quality translations, special translation methods have been developed and disseminated over the past decades.¹ Major international surveys (e.g., European Social Survey (ESS), European Values Study (EVS), Survey of Health, Ageing and Retirement in Europe (SHARE)) have been producing measurement instruments in English and their comparable translations, following best practice translation methods when possible.

The English source texts in these large-scale surveys and their translations are usually available online in a PDF format (see, for instance, ESS²) but not in a machine-readable format that can be automatically interpreted by a computer. Moreover, there is no information available in monolingual PDFs as to which translated elements correspond to which source text elements. To close this gap, the Multilingual Corpus of Survey Questionnaires (MCSQ), financed by the SSHOC project, was developed.³ This corpus allows the user not only to match the translated text of major surveys with the corresponding source text but also to download this “aligned” data in a machine-readable format that can be used to produce a Translation Memory (TM) for use within a computer-assisted translation (CAT) tool.

Computer-assisted translation environment and Translation Memory are two examples of “translation technology in a narrow sense,” which “includes only those tools that are (relatively) specific to the translation process.”⁴ A CAT environment is essentially an integrated suite of tools, with the Translation Memory being its main module.⁵ The Translation Memory itself is a repository of translated segments (sentences, sets of words, or single words) matched or “aligned” with their respective source text counterparts (see Figure 1 for an example of aligned text). A Translation Memory, when used in a CAT environment, allows translators to re-use previously translated segments by automatically proposing a relevant translation from the memory as a complete solution or partial solution whenever the same or a similar sentence occurs again in future work.⁶

¹ Harkness, “Questionnaire translation”; Behr and Shishido, “Translation of measurement instruments.”

² English source text, ESS: https://www.europeansocialsurvey.org/data/country.html?c=united_kingdom [May 2021]

³ Multilingual Corpus of Survey Questionnaires (MCSQ): <https://www.upf.edu/web/mcsq>; [May 2021]; Zavala-Rojas et al., “The Multilingual Corpus of Survey Questionnaires”

⁴ Krüger, “Contextualising computer-assisted translation tools,” 120f.

⁵ Bowker and Fisher, “Computer-aided translation.”

⁶ Reinke, “State of the art.”

WRITE IN DURATION:	NOTEZ LA DURÉE:
hours	heures
minutes	minutes
On a typical day, about how much time do you spend watching, reading or listening to news about politics and current affairs?	Au cours d'une journée ordinaire, combien de temps environ passez-vous à regarder, lire ou écouter les informations sur la politique et l'actualité?
Please give your answer in hours and minutes.	Veuillez donner votre réponse en heures et en minutes.
Refusal	Refus
Don't know	Ne sait pas
CARD 1	MONTREZ LISTE 1
People can use the internet on different devices such as computers, tablets and smartphones.	On peut accéder à Internet à partir de différents appareils tels que les ordinateurs, les tablettes et les smartphones.
How often do you use the internet on these or any other devices, whether for work or personal use?	Que ce soit sur ces appareils ou sur tout autre appareil, à quelle fréquence utilisez-vous Internet à titre professionnel ou personnel?
Never	Jamais
Only occasionally	Seulement de temps en temps
A few times a week	Quelques fois par semaine
Most days	Presque tous les jours
Every day	Tous les jours
Refusal	Refus
Don't know	Ne sait pas

FIGURE 1: SEGMENTED AND ALIGNED DATA FROM THE MCSQ (ENG_SOURCE_FRE_FR_ESS_R08_2016)

The underlying idea of these tools is to *assist* a human translator in the translation process. Contrary to machine translation (MT), where the computer translates the whole text, within a computer-aided translation environment and with Translation Memories, human translators translate a text by themselves by making use of computerized tools to help them effectively complete the task and increase their productivity.⁷

This report documents a use case scenario of Translation Memories for questionnaire translation within a CAT environment called MateCat. Moreover, a recorded demonstration for the use of Translation Memories accompanies this report.

2. MateCat and Translation Memory

MateCat⁸ is a freely accessible open-source CAT tool for the translation of texts, such as Word documents, Excel files, and many other file formats. It offers options for integration of different functionalities for the translation process, e.g., Translation Memories, machine translation, automatic spell-checking, etc. More

⁷ Bowker, *Computer-aided Translation Technology*.

⁸ "MateCat is an open source software released under the Lesser General Public License (LGPL) from the Free Software Foundation. Its open source license guarantees that you can download the software, integrate your own TM servers and machine translation engines" Source: Philosophy and Terms of service, MateCat; <https://site.matecat.com/terms/>; [May 2021]

information about the reasons for selecting MateCat for the current research project can be found in the *SSHOC Milestone 17. Open-source CAT TM software selected*.

One of the main functionalities of MateCat is a Translation Memory function. MateCat provides two different types of TMs: a generic public TM and a private TM.

The *generic public* translation memory called MyMemory is a large, publicly available source of Translation Memories, initially coming from European Union and United Nations projects. It can be accessed via an Application Programming Interface (API) by all MateCat users. It provides a search-like lookup of translated segments for a great number of combinations of source and target languages in various subject domains and is used extensively by MateCat. Moreover, it contains translations that translators using MateCat have voluntarily saved into this public memory. By default, MateCat uses MyMemory for its translations (see also *SSHOC Deliverable 4.7. Data exchange between TranslationCTRL and open-source CAT tools*).

The *private* TM consists of private projects that contain segmented translation units, with source text (original text for translation) and target text (translation) matched or “aligned” with each other. A private TM can be a newly created and completely empty database that is filled up with source and target text and updated in real-time during the translation process. The filling up and updating happens every time a segment is translated and/or approved in the tool.

On the other hand, a *private* TM can also be a newly created TM but with *imported* files containing previously translated text, which is uploaded before the actual translation process starts. To employ this option, the file with previously translated text needs to be in a special format called Translation Memory eXchange (TMX).⁹ This format was developed by a working group within the Localization Industry Standards Association and has been widely used in the professional translation community.

This latter type of private TMs is of key interest in this deliverable, since this report aims to showcase how the translations stored in the Multilingual Corpus of Survey Questionnaires can be of benefit for survey researchers in their translation activities.

Why should survey researchers use the Multilingual Corpus of Survey Questionnaires as the basis for Translation Memory? The use of the corpus as a TM can be beneficial for the following scenarios:

- Translators who wish to benefit from existing translations from major cross-national studies and to get ideas for their own translations.

⁹ TMX Format Specifications Version 1.0: <http://xml.coverpages.org/tmxSpec971212.html>; [May 2021]

- Translators who wish to replicate existing translations from the large-scale surveys in the corpus without having to search through PDFs (e.g., ESS Round 11 translators may wish to use ESS translations from Round 8). Please note, though: Section 6 lists what needs to be considered in order to have the metadata available that clearly specifies where (study, round, etc.) a translation in the TM came from.

And why should survey researchers use a TM at all, in particular a TM that is created during the translation process? The use of a TM can be beneficial in the following scenario:

- Translators who wish to ensure that there is consistency in the translation within a document. Consistency within a translation project should be checked to ensure the use of the same terminology, definitions, phrasing, scales, and the same wording within items' stem and answer options, etc.

3. Producing TMX files from the Multilingual Corpus of Survey Questionnaires

Before proceeding to the integration of TMs into MateCat, this section first explains how a TM can be obtained from the Multilingual Corpus of Survey Questionnaires. Users need to register for and log in to the corpus¹⁰. Under the heading "MCSQ datasets" they will then find the function "Create TMX." Here, they can select a target language (e.g. German) or a language-country combination (e.g., German from Germany), a study (e.g., ESS, EVS, SHARE) and/or year and then download a TMX file. The source language is always English because it is the source language of the surveys that are currently part of the corpus (ESS; EVS; and SHARE).

The resulting files will be called in accordance with filters selected for the download, e.g. "CAT_ESS_2002.tmx" (standing for Catalan – ESS study – year 2002). It is advisable to add the source language to the file name, namely "**ENG**_CAT_ESS_2002.tmx", so that at any time one can grasp from the title of the file what the language combination is that is contained in the TM, in this case: English – Catalan.

It is advisable to additionally download a csv file with the same language/study/year combination, again under the heading "MCSQ datasets" but this time using the function "Download alignment." The csv file will, unlike the TMX file, show for every segment where it came from (e.g., Source:

¹⁰ The Multilingual Corpus of Survey Questionnaires page: <https://www.upf.edu/web/mcsq/project> [July 2021]

"ESS_R01_2002_ENG_SOURCE_1016" and target "ESS_R01_2002_GER_AT_885"). Alternatively, you can open the downloaded TMX file with an XML viewer to see this additional information.

4. Integration of Translation Memories into a MateCat project and their use

In a short demo ("Use of Translation memories in survey translation") that can be accessed on Zenodo¹¹, the CAT tool MateCat and the integration of public and private Translation Memories are introduced. The recorded video includes the following steps:

- Adding the public TM to a project
- Adding multiple private TMs to a project, including the Multilingual Corpus of Survey Questionnaires
- Private TMs from other users / sharing private TMs
- Importing TMX files, exporting, and deleting TMs
- Using the TM key
- Using TM matches during the translation process: pre-fills, locking, exact and fuzzy matches, context matches, auto-propagation.

A few key terms from the demo shall be repeated here:

Matches in the TM can provide exact, in-context, or fuzzy information about previously translated text segments. While an *exact match* means that the entire segment of the source text for translation matches a source segment from a Translation Memory, an *in-context match* means that, beyond an exact match, the segments above and below are an exact match, too. *Fuzzy matches* include segments that have enough common content between a segment stored in the TM and the actual source text so that they can be useful as a starting point for translation. The threshold for usefulness must be defined by a translator during the translation task. Theoretically, a fuzzy match can have any value between a 99% and a 1% match. In practice, the default value for presenting fuzzy matches to the user in many CAT tools is between 65-70%; below this threshold, matches will not be shown unless a user changes the CAT tool settings.¹²

¹¹ Demo, "Use of Translation memories in survey translation": <https://doi.org/10.5281/zenodo.5102929> [July 2021]

¹² TM matches: <https://docs.sdl.com/783545/577385/sdl-trados-studio/tm-matches>; [May 2021]

During a translation project with several translators, a private TM can be used collaboratively if it is shared with others through a special private TM key. The special TM key is identified by a unique alphanumeric value. The key works as a password that enables collaborators to access the private TM. By using the private TM, the storage of translated and aligned text is visible exclusively for the creator of the TM and the person the key is shared with in contrast to the public TM, where every user can save their translations and overwrite the existing ones. For more information on how to share a private TM key with others, please refer to *MateCat Post-editing and outsourcing made easy*.¹³

5. Advantages and Disadvantages of Translation Memories in general

The use of Translation Memories always comes with positive and negative aspects that need to be taken into consideration when using Translation Memories for a translation project. Krüger defines the following advantages and disadvantages:

*For example, the use of TM systems may entail, on the positive side, a higher process standardisation, productivity and cross-document consistency and lower translation costs and, on the negative side, a possibly higher focus on individual translation segments (to the detriment of the textual perspective) or the uncritical adoption of low-quality translation units from contaminated TMs.*¹⁴

Other researchers (Christensen and Schjoldager) come to similar conclusions.¹⁵ The following positive and negative aspects can be summarized from their studies:

TABLE: ADVANTAGES AND DISADVANTAGES OF TM

Positive aspects	Negative aspects
translation with TM is experienced as easier	translators may lose control of the process
translation with TM is experienced as more interesting	translators may forget about the context

¹³ MateCat Post-editing and outsourcing made easy. User manual and installation guide:

https://site.matecat.com/wp-content/uploads/2014/12/MateCat-User-Manual-and-Installation-Guide_v1.6.3.pdf; [May 2021]

¹⁴ Krüger, "Contextualising computer-assisted translation tools," 115.

¹⁵ Christensen and Schjoldager, "Impact of Translation-Memory," 149.

translation with TM is experienced as faster and more efficient	translators may lose track of the text and the aim of the translation
translations become more consistent	translators may focus too much on the source text, the sentence level and terminology
the use of a TM allows the translator to easily make use of other people's knowledge	translators may lose their critical sense
	translating becomes more mechanical, less personal, less creative, less functional, and revision becomes more important

The usage of the **public TM - MyMemory**, which is suggested by default in MateCat means, on the one hand, that translators can use previous translations from a very large database for a great number of language pairs in various subject domains; furthermore, everyone translating in MateCat can contribute to this publicly accessible database and thus increase the volume of text stored in the public TM even more. On the other hand, the quality of these translations cannot be guaranteed. Thus, users of these **public TMs** should be careful when (partially) recycling existing translations.

6. Advantages and Disadvantages of Translation Memories from the Multilingual Corpus of Survey Questionnaires

Besides the advantages and disadvantages of Translation Memories in general, as listed above, using the Multilingual Corpus of Survey Questionnaires comes with its own conditions of use. When incorporating the TMX file provided by the MCSQ corpus into a private TM in MateCat, the following benefits and shortcomings/limitations need to be considered carefully.

Benefits include:

- Large database (appx. 657.000 segments) and hence ideas for one's own translations are provided.
- Appx. 80% of the data in the MCSQ corpus are "aligned" (i.e., source text segments and corresponding translations are matched) and can be exported in the TMX format used for import into a private TM in MateCat.
- The database consists of the most well-known international surveys in the social sciences domain.
- The translation process becomes more efficient if existing translations are displayed automatically in the translation field and can be (partially) recycled.

- [Studies using ESS, SHARE and EVS items] Translations for items existing in the corpus can be inserted automatically into the translation environment, and thus be replicated. This ensures consistency of existing translations from round to round in the same survey or across studies.
- The search functionality can be used for repeated questions to further strengthen the consistency between identical and similar questions.

Shortcomings/Limitations:

- The translation quality cannot be assured. Even in the ESS, known for its rigorous translation methodology, and where external translation quality checks have been incorporated in the translation process since its 5th round, consistently high quality of all translations cannot be guaranteed. The reliability of survey translations, even if already fielded, is always limited. Therefore, using already fielded questionnaire translations without quality check, even in well-known survey projects, is not recommended; careful case-by-case checking is always required.
- The Translation Memory itself or the corpus (using the corpus interface) do not provide information on whether translated items were pretested or validated or how they behaved in the actual survey. If one is interested in testing and validation procedures, documentation on translation from the actual surveys in the corpus needs to be accessed (if available).
- All surveys in the MCSQ corpus are face-to-face surveys and they are restricted to a certain set of topics relevant for social scientists. This means wording suitable for self-administered surveys as well as topics relevant for other disciplines will not be covered by the corpus.
- The suitability of translations and thus their transferability to new surveys depends on the respective questionnaire contexts; thus, even a 100% match provided by the TM may not work in the context of a new questionnaire due to different questionnaire contexts and survey needs.
- Not all segments in the downloadable TMX file provided by the MCSQ corpus are “aligned” (that means, in these cases a corresponding text in the other language does not exist) (but they do exist in appx. 80% of the entries).
- The alignment of the segments in the downloadable TMX file provided by the MCSQ corpus is done by an algorithm without human involvement¹⁶, and therefore the aligned segments can contain errors.

Some tips for practical TM use:

- The TMX file from the corpus, when implemented as a TM in a CAT tool, lacks metadata by design during the translation process, depending on how it was downloaded from the corpus. E.g., if one downloads *all* German language items from *all* studies of the corpus, the TM in the CAT tool does not show whether a suggested existing translation is from Germany, Austria, Switzerland, the ESS, EVS, or another study and from which round or wave from the study it came. Larger

¹⁶ Sorato & Zavala-Rojas, “Sentence Alignment.”

control of what is contained in the TMs requires that one downloads targeted TMs, e.g.: English – German (Germany) – ESS – Round 5 or English – German (Austria) – ESS – Round 5. The TMX download must be study-, round- and language-specific. If one then uses the same name for the TM in MateCat as the name of the downloaded TMX file, this name will also be visible in the CAT tool when using the TM and it will indicate to translators where a suggestion came from (in this illustration from either Austria or Germany – ESS – Round 5). See section 3 for more details of the download function.

- When using the MCSQ corpus as a TM, each translation project should clarify in advance how to use the corpus in case of similarity or identity of items with the TM. Should translations from the TM be adopted whenever possible – or should translators work with more leeway, making use of the MCSQ corpus TM as a source of inspiration, but beyond that take their own decisions on translation quality? One important question to ask in this context is as to whether any consistency and hence comparability to an existing survey (ESS, EVS, SHARE) is needed. Overall, though, the existence of Translation Memories will always call on discernment on the part of translators. Blind adoption should never be the rule.

7. Conclusion

This report presents computer-assisted translation technologies – the translation tool MateCat and Translation Memories – for use in questionnaire translation from a theoretical and practical perspective. The theoretical part contains definitions for CAT tools, for public and private TMs as well as their purpose and use scenarios. The practical part of this deliverable provides a recorded demo of the integration of TMs into a MateCat translation environment. The demo showcases how the public TM MyMemory and private TMs can be used and updated as well as how TM matches can be searched for during the translation process. The last part of the report describes the advantages and disadvantages of Translation Memories in general and of the public TM MyMemory as well as of TMs that can be created from the MCSQ corpus in particular.

Some known advantages of TMs include increased productivity and consistency, reduction of cognitive load, and possibly cost. Some known disadvantages are the lack of quality control of the previously translated texts that have been used to create some TMs¹⁷, and therefore uncritical usage of translations with low- or unknown quality, lack of or insufficient metadata, and the possibility of quick adoption of suggested translations without further search for a better alternative translation. The bottom line is that

¹⁷ In some translation companies and international organisations TMs are carefully reviewed, monitored, updated, and cleaned, so that their quality is high.

every translator needs to decide whether the benefits of using TMs outweigh the shortcomings, and they need to be aware of hidden traps of TMs.

While it is known that professional translators are increasingly using TMs in their work, not much is known about social scientists using TMs within computer-assisted translation environments when working on translations for international surveys. Therefore, to support social scientists in their increasingly digitized translation process, this report and the short demo of TM use have been produced. To increase adoption of and interaction with CAT technologies among social scientists, the development of more interactive formats is recommended, e.g., workshops, hands-on tutorials, or boot camps about translation technologies and their use for international survey translation projects. In the meantime, the official documentation on MateCat, as a case in point for a CAT tool, is a great source for learning even more in detail about the use of translation technologies, such as MateCat.¹⁸

Please note: Translation Memories and exported TMX files from the MCSQ corpus can be used in all types of CAT tools, beyond MateCat, and other CAT tools may have different functionalities compared to MateCat, for instance, additional search functions in a TM.

Please note: (Questionnaire) text that is confidential should not be translated in MateCat! Technical support from MateCat, for instance, would have access to the text uploaded to the MateCat environment for translation, in case technical problems need to be solved.

Please note: The use of a Translation Memory is not limited to using the download from the MCSQ corpus. Translation memories are built on the fly in every translation project that uses a CAT tool (see *private* TMs mentioned above). Furthermore, existing source questionnaires and their translations can be aligned in specific alignment tools by anyone and subsequently be used as a Translation Memory. Examples of free alignment tools include LF Aligner¹⁹ or the integrated alignment tool in MateCat²⁰. When building own Translation Memory resources, the restrictions to the surveys, to the source language English and to the language combinations provided by MCSQ corpus do not hold, of course.

This means: Researchers should take this deliverable as an idea of how Translation Memories, in general, can be used. The corpus serves as a powerful illustration, but your own resources can be built on top.

¹⁸ MateCat user guide: <https://guides.matecat.com/> [May 2021]

¹⁹ LF Aligner: <https://sourceforge.net/projects/aligner/> [July 2021]

²⁰ MateCat: <https://www.matecat.com/plugins/aligner/index#/> [July 2021]

8. References

- Behr, Dorothee, and Kuniaki Shishido. "The Translation of Measurement Instruments for Cross-Cultural Surveys." In *The SAGE Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-chi Fu, 269-287. London: Sage, 2016.
- Bowker, Lynn. *Computer-Aided Translation Technology: A Practical Introduction*. University of Ottawa Press, 2002.
- Christensen, Tina Paulsen, and Anne Schjoldager. "The Impact of Translation-Memory (TM) Technology on Cognitive Processes: Student-Translators' Retrospective Comments in an Online Questionnaire." In *Proceedings of the 8th International NLPCS Workshop: Special Theme: Human-Machine Interaction in Translation* (20.-21. August 2011), edited by Bernadette Sharp, Michael Zock, Michael Carl, and Arnt Lykke Jakobsen, 119–130. Copenhagen, Denmark, 2011.
- Harkness, Janet. "Questionnaire Translation." In *Cross-Cultural Survey Methods*, edited by Janet Harkness, Fons J. R. van de Vijver, and Peter Ph. Mohler, 35–56. Hoboken, NJ: Wiley, 2003.
- Krüger, Ralph. "Contextualising Computer-Assisted Translation Tools and Modelling their Usability." *Trans-Kom* 9, no. 1 (2016): 114–148. http://www.trans-kom.eu/bd09nr01/trans-kom_09_01_08_Krueger_CAT.20160705.pdf; [May 2021]
- Multilingual Corpus of Survey Questionnaires (MCSQ): <https://www.upf.edu/web/mcsq>; [May 2021]
- MateCat Post-editing and Outsourcing Made Easy. User Manual and Installation Guide. (Release 1.6.3.) https://site.matecat.com/wp-content/uploads/2014/12/MateCat-User-Manual-and-Installation-Guide_v1.6.3.pdf; [May 2021]
- MateCat User Guide: <https://guides.matecat.com/>; [May 2021]
- Pennings, Sebaastian. Deliverable 4.7. Code for Data Exchange Between TMT and Open-Source CAT Software, 2020. Zenodo doi: 10.5281/zenodo.4922730.
- Reinke, Uwe. "State of the Art in Translation Memory Technology." In *Language Technologies for a Multilingual Europe: TC3 III*, edited by Georg Rehm, Felix Sasaki, Daniel Stein, and Andreas Witt, 55–84. Berlin: Language Science Press., 2018. Zenodo doi: 10.5281/zenodo.1291930.
- Sorato, Danielly & Zavala-Rojas, Diana (Under review). Sentence Alignment of Bilingual Survey Texts using a Metadata-aware Strategy. Manuscript available upon request.
- Zavala-Rojas, Diana, Veronika Keck, Danielly Sorato, Dorothee Behr, and Brita Dorer. Report on Milestone 17. Open Source CAT TM Software Selected (Version v1.0), 2021. Zenodo doi: 10.5281/zenodo.4581806.

Zavala-Rojas, Diana, Danielly Sorato, Lidun Hareide, and Knut Hofland. The Multilingual Corpus of Survey Questionnaires: a tool for refining survey translation. *Meta: Journal Des Traducteurs*, forthcoming.