

The Multilingual Corpus of Survey Questionnaires (MCSQ)

Multilingual social surveys have become the main source of data to conduct cross-national comparative research. Translation procedures are transitioning to a larger adoption of Computer-Assisted Translation tools, which require the creation of text data.

The MCSQ corpus, the very first publicly available multilingual corpus of international survey texts, performs several functions. It enables the storing, searching and the comparison of information from international social surveys in 9 languages (e.g. French) and 29 of their language varieties (e.g. Swiss French).



The MCSQ constitutes an important new corpus for social scientists, linguists, translators and other interested parties. It has the potential to revolutionize survey research and design by allowing the researcher to consult both previous versions of surveys and also harmonize the same question across languages and language varieties. The Ask the same question method (ASQ) is at the heart of survey design. In the ASQ method, the translated questions in a multilingual survey are supposed to be functionally equivalent for the purpose of statistical analysis, that is, the data should be statistically comparable across linguistic groups. In the ASQ model, the concepts to be measured and the answer options are kept the same across languages, and cultural adaptation is only implemented to facilitate fluency and the use of locally appropriate terminology. Previously there has been no corpus available for ensuring that the same question is asked, and the MCSQ was compiled to fill this function.

Functionalities of the corpus:

- ➔ The MCSQ functions both as a repository of previous rounds of surveys and a corpus for systematic analysis of previous errors and successes.
- ➔ The corpus allows for the retrieval and preservation of source and translated questionnaires, and provides textual data for survey translation activities, research and training.
- ➔ The corpus facilitates the visualization and statistical analysis of previous translation decisions across languages. It is also possible to assess in a comparative perspective how a term or a collection of terms have been translated across different languages and in different contexts, and analyze retrospectively whether this decision was appropriate to communicate the intended source text message.
- ➔ The MCSQ also allows for the integration of translation analysis into the design of the source questionnaire.
- ➔ The corpus provides valuable training material for the highly specialized field of survey design and translation.

Benefits

The MCSQ is an open-access, open-source research and training resource. It is FAIR (Findable, Accessible, Interoperable and Reproducible) by design. The current version (version 3.0) named (Rosalind Franklin), is composed of 306 distinct questionnaires comprising approximately 766.000 sentences. Nearly 88% of the corpus is sentence aligned, meaning that the source text sentences in English are linked to their translations in the different languages. In total, it is compiled from the following surveys and their versions:

- ➔ European Social Survey (ESS): Round 1, Round 2, Round 3, Round 4, Round 5, Round 6, Round 7, Round 8, Round 9
- ➔ Survey of Health, Ageing and Retirement in Europe (SHARE) Round 7, Round 8 and COVID-19 questionnaire
- ➔ European Values Study (EVS): Wave 2, Wave 3, Wave 4, Wave 5
- ➔ WageIndicator: Round 1 and COVID-19 questionnaire

The surveys are available in English language and their translations into different languages (the availability of the language varies from the survey and its version): Catalan, Czech, French (localized language varieties for France, Switzerland, Belgium and Luxembourg), German (localized for Austria, Germany, Switzerland and Luxembourg), Norwegian (Bokmål), Portuguese (localized for Portugal), Spanish (localized for Spain) and Russian (localized for Belarus, Estonia, Israel, Latvia, Lithuania, Russia and Ukraine).

More Information

- ➔ <https://www.upf.edu/web/mcsq/>
- ➔ [Manual; Presentation](#)
- ➔ [SSHOC Deliverable 4.3 Survey specific parallel corpora](#)
- ➔ [MCSQ Version v2.0 \(Mileva Marić-Einstein\)](#)
- ➔ [Blogpost](#)
- ➔ [MCSQ version 3.0 \(Rosalind Franklin\)](#)
- ➔ [The Multilingual Corpus of Survey Questionnaires \(MCSQ\) at European Survey Research Association \(ESRA\) Conference 2021](#)

Who made the tool

Diana Zavala-Rojas (European Social Survey ERIC, Universitat Pompeu Fabra), Danielly Sorato (Universitat Pompeu Fabra), Lidun Hareide (Moreforskning Research Institute in Norway), Knut Hofland (University of Bergen). The use of the MCSQ data is licensed under [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#). The use of the MCSQ software and source code is licensed under EUPL-1.2 ([European Union Public License](#)).

Access the MCSQ

Access the MCSQ



Access via the SSHOC Service Catalogue



Datasets

Multilingual
Corpus of Survey
Questionnaires



Dataset