What to consider and look out for in questionnaire translation

Dorothée Behr

Abstract

The recommendations in this guideline focus on the concrete translation step, answering the question what those translating questionnaires need to keep in mind and consider when translating and taking actual translation decisions. Areas covered include: (a) the translation of meaning, (b) cultural adequacy, (c) response scales, d) consistency, e) style and register, f) linguistic correctness, and g) layout and presentation. We particularly have questionnaire translations for cross-cultural or cross-national surveys in mind where comparability is paramount.

Citation

1. Focus of this guideline and reference to other guidelines

This document focuses on the **concrete translation step**, answering the question what those translating questionnaires need to keep in mind and consider when translating and taking actual translation decisions.

Guidelines notably focusing on the ***translation procedure*** (incl. parallel translation and team-based review, the so-called TRAPD model) can be found in Mohler, Dorer, de Jong, & Hu (2016) as well as in Behr, Braun, & Dorer (2016).

Detailed information on ***documentation*** aspects of questionnaire translation can be found in Behr and Zabal (2020).

Furthermore, the European Social Survey (ESS) provides both guidelines on the translation procedure as well as on content related-aspects of translations (European Social Survey, 2020).

2. Terminology

In the following, "source X" will be used to indicate the original questionnaire, language, culture, and country. "Target X" will be used to denote the translated text (as in "target questionnaire"), the language of translation as well as the country or culture of the target population.

3. Translation brief

Before embarking on a questionnaire translation or before commissioning someone else to do a questionnaire translation, the goal of a specific questionnaire translation needs to be specified, alongside the requirements that go with this goal. Getting it right for oneself or informing someone else on what needs to be heeded in translation often happens through a so-called *translation brief*. A good translation meets the requirements of a specific translation brief (which can be different between survey projects). Here is a list that *should* (marked by *) or *can* be part of such a brief:

- Purpose of a translation: *
    - There are mainly two purposes of questionnaire translations: a) producing a comparable measurement instrument in the target culture and b) producing a 'documentary translation' that 'documents' what is asked in a given country/culture without adapting it to a new context; the latter is often the case for archives that contain foreign language questionnaires that external researchers would not understand without such a 'documentary translation.'
- Target population: *
    - Description of target country, target language, age, education, native tongue, profession or other important socio-demographic criteria; French, for instance, without further reference to the country of use is insufficient to describe the target population
- Survey mode: *

- o Indication of survey mode is crucial since formulations may differ depending on whether a text will be administered by an interviewer or read by the respondents themselves; whether respondents only hear text or whether they can see it themselves, including visual clues; also web administration may differ from paper administration (e.g. related to instructions).
- Source questionnaire origin: *
    - o Questionnaires that have been specifically designed for a cross-cultural or cross-national study often call for translations (as opposed to adaptations, i.e., intentional deviations to make a questionnaire suitable for a new target culture). Questionnaires that were originally developed in another country and are now supposed to be used in a new cultural setting may require (a larger amount of) adaptations that will have to be identified in a thorough translation and adaptation process.
- Project-specific guidelines that need to be heeded (if available):
    - o May include information on how to deal with cultural challenges (e.g. when the source instrument does not fit the target culture)
    - o May include information on the desired translation approach
    - o May include Information on how to deal with existing translations (if available from previous waves, for instance)
- Reference material: glossaries, style guides, reference studies, concept elaboration of items, etc. (if available)

4. Translation requirements *in a nutshell*

In a nutshell, questionnaire translations used for cross-cultural or cross-national research should

- retain the meaning and the measurement properties of the source questionnaire (*comparability*);
- adhere to general questionnaire design requirements, such as clarity, fluency or typical text-type convention for the text genre questionnaire; adhere to the grammatical, idiomatic, and pragmatic requirements of the target culture.

5. Translation requirements *in detail*

In the following, we will raise awareness of aspects that need to be heeded and/or potential challenges in questionnaire translation. **This is a very general overview; in reality, exceptions and trade-off decisions are likely necessary. Translation is always an individual case-by-case decision, and, moreover, it depends on the specific language pairs and contexts: What works in one language, may not work in another language.** The more the target language structure (syntax, grammar, etc.) is apart from the source language structure, the more deviations are likely needed.

5.1 Meaning

Conveying the meaning of the source questionnaire is essential; misrepresentations of meaning can take on the following forms:

- (Obvious) misinterpretations
- Shifts in meaning (e.g., formulation is too narrow or too wide)
- Unsuitable connotations (i.e., unwanted associations triggered by terms in the target language)
- Unclear meaning
- Unintended ambiguity
- Potential misunderstanding
- Omission of meaningful text elements (e.g., "in general", "in total")

When translating and/or reviewing a questionnaire, these aspects should be checked.

Please note: If the source questionnaire is already problematic (e.g., the meaning of a term, such as 'upset', is unclear in a given context, the developers of the source questionnaires should be approached to clarify the meaning).

Table 1: Examples

Obvious misinterpretations: "Wealthy" was translated as "bonne santé" [healthy] (Fitzgerald, Widdop, Gray, & Collins, 2011).

Shifts in meaning: For the item "If people who have come to live here commit <u>any</u> crime, they should be made to leave," the chosen Danish translation ("lovovertrædelse") roughly translates back as 'breach of the law' – this translation had a negative impact on comparative measurement (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014, and Fitzgerald et al., 2011).

Potential misunderstanding in the translation: "Now think about your best friend…" In many languages, "friends" will need to be translated by explicitly referring to both male and female friends to suggest to respondents that all friends are meant (Haas, 2009).

Tip 1: Monolingual dictionaries can help to identify different meaning dimensions of a term and single out the one that best fits in a given context. A great online source (and entry portal) for numerous English-language monolingual dictionaries is: https://www.onelook.com/.

Tip 2: In the table above, one example refers to gender. This topic is increasingly discussed under the term "gender-sensitive" language. Given the evolutions in many societies, the topic of gender-sensitive language should be integrated into questionnaire translations. The topic is relevant for: (1) addressing respondents; (2) all items that include a term indicating a person (e.g., employee, nurse, someone, friend, partner). The solutions need to take into account the measurement goal of items, what is currently (politically) correct in a country, acceptability among the target population, the usual handling of gender-sensitive language by the survey agency, and ultimately also the ease or complexity of survey items.

5.2 Cultural adequacy

Adaptations should be integrated into questionnaire translation if the content of the source questionnaire conflicts with the cultural or socio-political background of the target country.

Content that may require adaptations can include traditions, customs, behavior, pre-supposed knowledge, references to the political or economic system, etc. See the appendix for an overview of adaptation types and topics.

Questionnaires that have specifically been developed for a cross-cultural survey typically avoid or, of this is not possible, earmark items for adaptations. Questionnaires that have been developed with a specific culture in mind are prone to adaptation needs when the instrument gets transferred to a new cultural context.

Table 2: Examples

> **Cultural adaptation**: "family investing in the stock market" (US English). For a Mexican version, this item was changed into "family saving in certificate deposits", since investment in the stock market was not part of the usual financial practice in Mexico. (Arce-Ferrer, 2006)
>
> **Cultural adaptation**: The item "I am concerned that my medicines interact with alcohol" was adapted for an Arabic-speaking population as "I am concerned that my medicines interact with my nutritional habits (other foods, alcohol, drinks)" to take into account that alcohol consumption is not the norm in the Arabic culture. (Zidan, Awaisu, Hasan, & Kheir, 2016)

5.3 Response scales

Response scales are at the heart of measurement since here the answers are recorded that later lead to the statistics. Response scales (now focusing on ordinal and interval scales) typically have a *dimension* that is measured (e.g., satisfaction, agreement, happiness, etc.) and a *quantification/negation* (e.g., strongly, very, not, dis-). Both components should ideally be matched in the translation (Harkness, 2003).

In general, we would expect the following from a questionnaire translation:

- Semantic comparability to the source scale, that is, paying attention to the *dimension* and the *quantification/negation* of the scale., (e.g., endpoints should not be more or less extreme than in the source).
- Retaining further scale properties such as the unipolar vs. bipolar nature of a scale.
- Ensuring equal distance between scale points if existing in the source scale.
- Ensuring linguistic symmetry of wording if existing in the source scale.
- Disjunct response categories (= excluding each other, especially relevant for numerical scale points).
- Idiomatic and linguistic fitting between adverbs (very, quite, etc.) and the corresponding adjectives (satisfied, happy, etc.).
- (Grammatical) fit between question and scale.

Beyond the response scales themselves, care should be paid to the question wording itself. Balanced wording (e.g., "To what extent do you agree or disagree …"? or "How easy or difficult …") is a way to signal to respondents that both response directions, those agreeing and those disagreeing, are totally acceptable. Moreover, words such as "To what extent …" or "How …" indicate that a range of response options will be presented. Simply asking "Do you agree or disagree that …" would rather indicate that a yes/no answer is expected.

The scale translation should also be considered together the translation of an item; double negation across item and scale (if not present in the source) should be avoided to reduce misunderstanding and the potential for bias.

Table 3: Examples

> **Semantic comparability – extreme endpoints**: For a scale running from "extremely dissatisfied" to "extremely satisfied" the endpoints, where nothing can go beyond, should be translated in this extreme manner; the exact words could be 'extremely', 'fully', 'completely', etc., and depend on the respective target language.
>
> **Unipolarity vs. bipolarity:** For a scale running from "good" to "bad" (i.e., a bipolar scale), we would equally expect a bipolar scale in translation rather than a unipolar scale running from "good" to "not good."
>
> **Linguistic symmetry of wording:** For the example given above, we would expect the same adverb ("extremely") on both ends of the translated scale.
>
> **Disjunct response categories**: "(1) Less than ½ hour; (2) ½ hour to 1 hour; (3) more than 1 hour, up to 1 ½ hours; etc." This source scale is very explicit in that it guides the respondents clearly to unambiguous response categories. A translation needs to match the same differentiation and not, for instance, formulate the third category as "1 – 1 ½ hours." In this case, the answer "1 hour" could equally fit to category (2) or (3), leading to measurement error.

However, while the expectations are clear, there may be reasons for diverting from "close" translations of a source response scale; these may be linguistic necessities or constraints of the target language (e.g., linguistic symmetry cannot be maintained; or a bipolar scale needs to be rendered as a unipolar scale) or intentional decisions to give preference to home-grown scales or scales more appropriate for the response style of a target culture (see also Harkness, 2003).

The best recommendation in such cases is to document decisions and deviations so that these may be taken into account in analyses.

Tip 3: To receive inspiration from other surveys' translations, **the Multilingual Corpus of Survey Questionnaires** can be consulted and easily accessed with various search functions: https://www.upf.edu/web/mcsq/project. The MCSQ contains in its current version (Version 3) the English source and its translations into Catalan, Czech, French, German, Norwegian, Portuguese, Spanish, and Russian for 306 questionnaires comprising 766.000 sentences. The following surveys are covered: European Social Survey (ESS), the European Values Study (EVS), the Survey of Health Ageing and Retirement in Europe (SHARE), and the Wage Indicator (WI) survey. However, the pure existence of a scale translation should not be taken as evidence that this scale works in a comparable fashion in statical terms. But the versions offered in the corpus can certainly be a source of inspiration.

5.4 Consistency

Consistency is crucial for a consistent stimulus and reduced respondent burden. Consistency needs to be ensured and checked on various levels:

- Consistency in item wording between different waves of a study (replication);
- consistency of repeated elements in a questionnaire, e.g.
  - interviewer/respondent instructions,
  - core concepts,
  - response scales (e.g., reoccurring response scales),
  - almost identical questions,

o   response scale and the labels of the response scale – if these are already included in the question text.

Table 4: Examples

Consistency of almost identical questions: The PIAAC questionnaire[1] includes a section for respondents with a current job (e.g., "*In this job, are you working as an employee or are you self-employed?*") and a section for respondents with a past job (e.g., "*In this job, were you working as an employee or were you self-employed?*"). With the exception of the tense (are/were) these questions should be translated consistently.

Consistency between question text and response scale: If a response scale is referred to first in the question itself and then as a response scale, the response categories should be translated consistently between these two places (ESS, R6).[2]
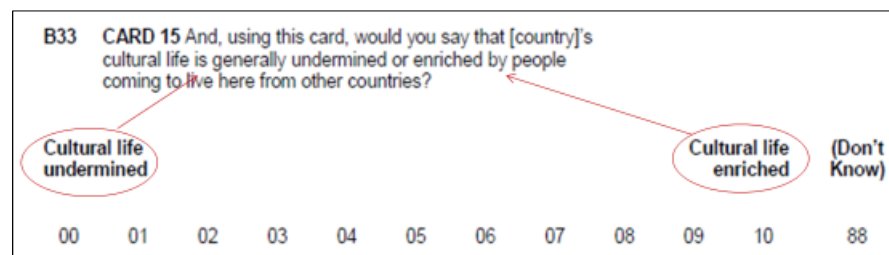
A note of caution: Translation is context-bound; it may very well be that the source language uses a term or response scale consistently throughout, but that in the target language different wordings will be needed. Consistency should therefore always be decided based on the given context.

Tip: In the translation industry and in larger cross-national surveys, computer-aided translation tools (CAT tools) are frequently used. These computer environments support the translation process, for instance, by automatically proposing (partial) solutions that have been produced before for an identical or similar source text segment. This function is called Translation Memory (TM); further details on TM use in questionnaire translation can be found in Keck, Behr, and Dorer (2020).

5.5 Style and register

Using an appropriate style and register (the latter meaning the style of language, grammar, words used for a particular situation) is crucial for correctly conveying the intended meaning. The following aspects should guide questionnaire translators:

- Short, simple and clear sentence structure;
- idiomatic wording (i.e., which sounds natural in the target language and does not come across as a translation);
- simple wording;

---

[1] https://www.oecd.org/skills/piaac/Background%20Questionnaire%2015DEC10.pdf (18 October, 2021)

[2] https://www.europeansocialsurvey.org/docs/round6/fieldwork/source/ESS6_source_main_questionnaire.pdf (24 February 2023)

- adequacy in terms of typical question formulations and questionnaire conventions (e. g, when addressing men/women if languages differentiate in this regard);
- adequacy of the language used for the target population considering age, education and other relevant socio-demographic characteristics;
- adequacy of the translation for the mode (*oral* communication face-to-face or by telephone, *written* communication paper-based or computer-based).
- complying with cultural conversational conventions (politeness, form of addressing, etc.)

These aspects apply to questionnaire developers in monolingual contexts, too. In general, for (frequent) questionnaire translators it makes sense to familiarize themselves with general questionnaire design principles so that they can recognize these easily in questionnaire translations.

Table 5: Examples

Adequacy of the language used for the target population: While the English language only knows the pronoun "you" for addressing persons, other languages will have to distinguish between, for instance, "Sie" or "vous" for adults and "du" or "tu" for children or for less formal contexts.

Adequacy of the translation for the mode: A survey that will be read out by an interviewer may need to be worded differently than a survey that can be read on the screen. Mode also influences other decisions, such as instructions for selecting response options (e.g., click on, check, mark with a cross, etc.)

Complying with cultural conversational conventions: "Would you please be so kind as to give me your e-mail address so that we can send you the online version of the questionnaire?" Here, the polite request needs to be formulated in a way that corresponds to the way how such requests are typically worded in a speech community. Literal translations, without taking into account language use in a country, would certainly not be appropriate. (Behr, Dept, & Krajčeva, 2018).

Computer-assisted surveys come with additional constraints for translations, in particular if placeholders are integrated into surveys (also often referred to as 'fills'). Here, both an understanding of the role of placeholders as well as close collaboration between survey developers and programmers on the one side and translators on the other side are needed; after all, programming that works in one language does not necessarily work in other languages (Behr, 2022).

Tip 4: Especially for interviewer-administered surveys we recommend reading out loud the translations, when producing them (e.g., during team-based discussions), in order to check whether questions could be smoothly delivered by an interviewer without introducing measurement error.

5.6 Linguistic correctness

A questionnaire translation should be linguistically correct. Errors in this regard undermine the seriousness of a study. For this, the following aspects should be borne in mind and checked:

- Spelling;

- punctuation;
- syntax and grammar;
- collocation (= typical combinations of words).

In particular for syntax and grammar it should be checked whether it is natural and correct. Especially those not used to and familiar with translation stick too closely to source text structures and thus neglect the needs of the target language. The resulting translation may be unidiomatic, awkward or, in the worst case, not intelligible.

Table 6: Examples

Syntax and grammar: "Using this card, would you say that …". The English language gerund "using this card…" likely needs a rewording in other languages that do not use the gerund in such ways. A translation may end up saying something along the following lines: 'Please use this card and tell us ….'.

Collocation: In particular when it comes to response scale categories combining adverb (very, quite, rather….) with adjective (satisfied, efficient, etc.) or verb (agree, approve, etc.), the resulting word combination should 'collocate' and fit together.

5.7 Layout/presentation

Besides text, also the layout and presentation of comparative surveys should be equivalent. This applies to:

- Equivalent text formatting (bold, underlined, etc.);
- layout of the response scales (horizontal, vertical, etc.).

Translated text may be longer in translation, which may mean that it no longer fits into the original layout (e.g., paper layout/web layout). Thus, the translation should carefully be checked and adjusted in the final survey mode and layout (e.g., Behr, 2022, for checks in translated web surveys).

Table 7: Examples

Equivalent text formatting: Sometimes text within an item may be underlined to stress a particular element. E,g., "Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out" (ESS, R9). This emphasis should be maintained in the translations.

Layout of the response scales: The scale direction (horizontal, vertical) or format (pyramid, ladder, etc.) influence measurement. These characteristics should be maintained in the translations.

6. Further reading

Pan, Y., Sha, M., & Park, H. (2019). *The sociolinguistics of survey translation*. Routledge.

Bowker, L. (2023). *De-mystifying Translation: Introducing Translation to Non-translators*. Taylor & Francis. https://www.taylorfrancis.com/books/oa-mono/10.4324/9781003217718/de-mystifying-translation-lynne-bowker

7. Appendix

7.1 Cultural adaptations

Cultural adaptations, i.e. intentional deviations to make a question more suitable for a new target culture

Table 8: Examples of adaptation types and topics (Harkness, Villar, & Edwards, 2010, p. 135).

Topic	Details	Concrete Examples
Measurements	Distance; temperature; weight; length/height	100 yards; 20ºC; 2lb. (imperial); 6ft
Components and order of calendar dates, names, initials, addresses	Order of day/month/year; order of first and last names and initials of these; components of addresses and their order	13/1/2011; Harkness, Janet; HJ (Hungarian initials order); street number, apartment/house number, zip code, place name, regional division, country
Institutions	Schools; institutionalized services and practices; government bodies	Grammar school; VAT; social security; welfare benefits; Parliament
Product & food terms, brand names	Cigarettes, cooking oil, vegetables and fruits	Light-tar cigarettes; names of products
Sexual terminology	Sexual orientations, sex practices, parts of the body	
Emotions and conditions	Instances and intensities of feelings or states	Anxiety; "brain fog"
Socio-economic questions	Education; income; type of housing	(See Chapter 17, this volume)
Answer scales	Endpoints; midpoints; other points	(See Chapter 3, this volume and this chapter)
Visual images	Saliency, focus, processing habits; semiotics	Perceived foreground & background; direction of reading/ processing; socially acceptable depictions.
Direct and indirect disclosure formats	Habitual indirect disclosure; questions sensitive only in some populations	Names of children
Cultural and social conceptions and conventions	An extremely wide range of topics could be involved	Hygiene practices; taboos, health care

7.2 Machine translation errors

Machine translation (MT) can nowadays be reached by everyone at a finger click. Professional translators often use it as a basis for further revisions; the activity of using MT and revising it is called *post-editing* in the translation industry (Nitzke & Hansen-Schirra, 2021).

In the following, types of MT errors are outlined to raise awareness of the fact that raw MT output typically requires further checking and revision steps. Translators may use MT to produce a first draft of a questionnaire translation. However, team-based translation approaches such as recommended in Behr, Braun, and Dorer (2016) or Harkness (2003) should still follow. First-of-its-kind research of integrating MT into a comprehensive translation and assessment approach to questionnaire translation can be found in Zavala-Rojas et al. (forthcoming).

Neural machine translation, the currently dominating MT paradigm (e.g., which DeepL Translate and Google Translate are based on) may lead to quite fluent MT output. This, in turn, can lead to overlooking errors (Nitzke & Hansen-Schirra, 2021). Neural MT quality is driven mostly by mistranslations and omissions (Moorkens, Toral, Castilho, & Way, 2018), which, if undetected, could be detrimental for survey comparability. MT errors do partly overlap with typical human translation errors and partly diverge from those, which is why MT should not be taken lightly.

Important: The quality of MT systems is dependent on language pair, domain, and text type and it also hinges on whether or not a MT system has been trained for a specific domain or text type. Moreover, free online MT systems (e.g., DeepL Translate and Google Translate) should not be used for confidential survey material because the source text is often saved on the provider's servers and so it might become accessible to third parties (Nitzke & Hansen-Schirra, 2021).

Examples of MT errors: [3]

- **(Unexpected) mistranslations or omissions**
    - *"…they turn down a job because it pays a lot less than they earned previously*?" In German, this became "*… Sie einen Job ablehnen, weil er sich viel weniger auszahlt als zuvor?*" "*They*" got translated as "*Sie*," which can be read as the formal personal address (due to the capital 'S') and thus the MT translation incorporates the respondents themselves into the question.
    - "*How often have you…*" was translated as "*Wie oft hast du …*", with "*you*" translated in the informal address "*du*," which is used only for children or persons that are known, such as family and close friends, but not for respondents in a social science survey targeting the general population.
    - "*Could not take the time off work*" translated as "*Ich konnte mir keine Zeit für die Arbeit nehmen*" [[I could not take time for work], thus reversing the original meaning.
    - "*…lead to a more equal society?*" translated as "* … Zu einer gleichberechtigten Gesellschaft führen?*" [… leading to an equal society], thus omitting the comparative.
- **MT not taking into account the need for gender-inclusive language**
    - "*Imagine someone in their 50s who is unemployed and looking for work*" (Translation instruction: 'Their' meaning a gender neutral term for a single person.) In German, this became "*Stellen Sie sich jemanden in den Fünfzigern*

---

[3] The examples are all taken from an MT project within the EU project SSHOC (Zavala-Rojas et al., forthcoming).

vor, der arbeitslos ist und Arbeit sucht." [Imagine someone in the 50s who (male pronoun) is …].

- **MT not taking into account survey-specific language, such as (context-deprived) scale labels**
  - "*Strongly agree – agree – neither agree nor disagree –disagree – strongly disagree*" scale in English: In German, the MT output used three different translations for "disagree" within one scale (*abstreiten – nicht zustimmen – widersprechen*), which in two instances are even far from appropriate for the survey context.
  - "*Agree with statement A – Agree with Statement B*"; in German, these response labels were translated as a question rather than a declarative sentence ("*Stimmen Sie der Aussage zu*" [Do you agree with statement A])
- **MT not taking into account consistency**
  - "*For each activity, would you say you do them every week or nearly every week, once or twice a month, only a few times a year, or not at all*?" "*Spend time with friends*," "*Spend time with people at your church, mosque or synagogue*." In German, while the first "Spend time with…" was syntactically translated correctly, the second "Spend time with …" was turned into a question ("*Verbringen Sie Zeit mit Menschen in Ihrer Kirche, Moschee oder Synagoge*" [Do you spend time with….].
- **MT leading to overly literal translations**
  - "*Now I'd like you to look at this card*." In German, conversational elements such as these were often translated on a word-by-word basis ("*Jetzt möchte ich, dass Sie sich diese Karte ansehen*" [Now I want that you look at this card], leading to awkward wording that is unsuitable for the survey context.
- **MT not being able to cope with "broken stem questions,"** which distribute information across several linguistic segments, such as "*Using this card please tell me where most people would place the status of…*" "*…people in their 20s*?" In German, the verb was omitted altogether, which shows that there can be partly unexpected omissions, which are not among typical human translation mistakes.
- **MT not being able to anticipate register (i.e., style of language, words) of the target group**
  - "*The person from a different race or ethnic group than most [country] people is more likely to be found guilty*." "*Ethnic*" was translated as "*ethnisch*," which is high level register and likely not understood by many in a general population survey.
- **MT not being able to cope with adaptations**
  - "*I'm interested in how you think most people in [country] view the status of people in their 20s, people in their 40s and people over 70*." MT does not substitute the target country name for "*country*."
  - "*Junior level non-manual - office worker etc*." Jobs, occupations are usually challenging to translate since these often need country-specific wording. The word-by-word translation "*Junior manuell nicht manuell - Büroangestellte etc*." is nonsense.

7.3 Summary: Evaluation grid – Questionnaire translations (English)

Evaluation grid[4]

> *Note: Depending on language combination and survey context it may not be possible to meet all the provided dimensions; rather it may be necessary to make trade-off-decisions.*

Checking of meaning

Possible sources of error:
- (Obvious) misinterpretations
- Shifts in meaning (e.g., formulation is too narrow or too wide)
- Unsuitable connotations
- Unclear meaning
- Unintended ambiguity
- Potential misunderstanding
- Omitting of meaningful text elements (e.g., "in general", "in total")

Checking of cultural adequacy

- Concerning the cultural background (behavior, traditions, customs, relevant knowledge, …) → **cultural adaption necessary?**
- Concerning the socio-political background (political system, economic system, etc.) → **cultural adaption necessary?**

Checking of style and register

- Short, simple and clear sentence structure
- Idiomatic wording
- Simple wording
- Adequacy in terms of typical question formulations and questionnaire conventions (e. g, when addressing men/women if languages differentiate in this regard)
- Adequacy of the language used for the target population considering age, education and other relevant socio-demographic characteristics
- Adequacy of the translation for the mode (*oral* communication face-to-face or by telephone, *written* communication paper-based or computer-based).
- Complying with cultural conversational conventions (politeness, form of addressing, etc.)

Checking of response scales

- Semantic comparability to the original, that is, paying attention to the *dimension* and the *quantification*/*negation* of the scale
- Retaining further scale properties such as the unipolar vs. bipolar nature of a scale.

---

[4] Based on Behr (2009), see also Pan und de la Puente (2005)

- Ensuring equal distance between scale points if existing in the source scale.
- Ensuring linguistic symmetry of wording if existing in the source scale.
- Disjunct response categories (= excluding each other, especially relevant for numerical scale points).
- Idiomatic and linguistic fitting between adverbs (very, quite, etc.) and the corresponding adjectives (satisfied, happy, etc.).
- (Grammatical) fit between question and scale.

Checking of linguistic correctness

- Spelling
- Punctuation
- Syntax and grammar
- Collocation (= typical combinations of words)

Checking of consistency

- Consistency in item wording between different waves of a study (replication);
- consistency of repeated elements in a questionnaire, e.g.
    - interviewer/respondent Instructions,
    - core concepts,
    - response scales (e.g., reoccurring response scales),
    - almost identical questions,
    - response scale and the labels of the response scale – if these are already included in the question text.

Checking of layout/ presentation

- Equivalent text formatting (bold, underlined, etc.)
- Taking over the layout of the scale (horizontal, vertical, etc.)
- (If applicable) Cultural adaptations (colors, text boxes, writing direction, etc.)

**Evaluationsraster –
Hilfestellung zur Erstellung oder Evaluation von Fragebogenübersetzungen[5]**

> *Hinweis: Je nach Sprachkombination und Survey-Kontext können möglicherweise nicht
> alle dieser „Richtlinien" eingehalten werden bzw. es werden trade-off-Entscheidungen
> gefällt werden müssen.*

Überprüfung der Bedeutung

Mögliche Fehlerquellen:

- (Klare) Fehlinterpretationen
- Bedeutungsverschiebungen (z.B. zu eng oder weit formuliert)
- Unpassende Konnotationen
- Unklare Bedeutung
- Nicht beabsichtigte Mehrdeutigkeit
- Potenzial für Missverständnisse
- Auslassung bedeutungstragender Textelemente (z.B. ‚in general', ‚in total')

Überprüfung der kulturellen Adäquatheit

- vor dem kulturellen Hintergrund (Verhalten, Traditionen, Sitten, Wissensbestände…) →
  kulturelle Adaptation notwendig?
- vor dem sozio-politischen Hintergrund (politisches System, wirtschaftliches System
  etc.) → **kulturelle Adaptation notwendig?**

Überprüfung von Stil und Register

- Kurzer, einfacher und klarer Satzbau
- Idiomatische Sprache
- Einfache Worte und Formulierungen
- Adäquatheit in Bezug auf typische Frage-Formulierungen und Fragebogen-
  Konventionen (z.B., bei Ansprache von Männern/Frauen, wenn Sprachen hier
  Unterscheidungen machen)
- Adäquatheit der Sprache für die Zielgruppe unter Berücksichtigung von Alter, Bildung
  und sonstigen relevanten sozio-demographischen Merkmalen
- Adäquatheit der Übersetzung für den Modus (*mündliche* Kommunikation face-to-face
  oder telefonisch, *schriftliche* Kommunikation papier-basiert oder computer-basiert)
- Einhaltung kultureller Sprachkonventionen (Höflichkeit, Anrede, Sprechakte usw.)

Überprüfung von Antwortskalen

---

[5] Basierend auf Behr (2009), siehe auch Pan und de la Puente (2005)

- Semantische Vergleichbarkeit zum Original, d.h. Beachtung der *Dimension* und der *Quantifikation*/*Negierung*, die in der Skala zum Ausdruck kommt
- Wahrung von Skaleneigenschaften, wie unipolare vs. bipolare Skalen
- Symmetrie zwischen den Skalenoptionen, sofern im Original vorhanden
- Gleichabständigkeit zwischen den Skalenoptionen, sofern im Original vorhanden
- Disjunkte Antwortkategorien (= sich gegenseitig ausschließend, insbesondere relevant für numerische Skalenpunkte)
- Idiomatische und sprachliche Passung zwischen modalen Adverbien (sehr, ziemlich usw.) und dazugehörigem Adverb/Adjektiv (zufrieden, glücklich usw.)
- (Grammatikalische) Übereinstimmung zwischen Frage und Skala

Überprüfung der sprachlichen Korrektheit

- Rechtschreibung
- Zeichensetzung
- Syntax und Grammatik
- Kollokationen (= typische Wortkombinationen)

Überprüfung der Konsistenz

- Konsistenz zwischen verschiedenen Wellen einer Studie (Replizierung)
- Konsistenz innerhalb eines Fragebogens von wiederkehrenden Elementen, z.B.
  - Instruktionen
  - Kernbegriffen
  - Skalen
  - Fragen, die nahezu identisch sind
  - Skala, vorweggenommen in Fragetext, und entsprechende Antwortskala

Überprüfung des Layouts/der Präsentation

- Äquivalente Textformatierung (fett, unterstrichen usw.)
- Übernahme des Skalenlayouts
- Ggf. kulturelle Anpassungen (Farben, Textfelder, Sprachrichtung etc.)

References

Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme-response style: Improving meaning of translated and culturally adapted rating scales. *Educational and Psychological Measurement*, *66*(3), 374-392. doi: 10.1177/0013164405278575

Behr, D. (2009). Translationswissenschaft und international vergleichende Umfrageforschung. Qualitätssicherung bei Fragebogenübersetzungen als Gegenstand einer Prozessanalyse. Bonn. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-282120

Behr, D. (2022). Computer-assisted migration research: What we can learn from the field of software localization for source questionnaire design and translation. In S. Pötzschke & S. Rinken (Eds.), *Quantitative Migration Research in a Digitized World. Using Innovative Technology to Tackle Methodological Challenges*. Springer.

Behr, D., Braun, M., & Dorer, B. (2016). Measurement instruments in international surveys. *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_006

Behr, D., & Zabal, A. (2020). Documenting survey translation. Mannheim, Germany: GESIS – Leibniz-Institute for the Social Sciences (GESIS – Survey Guidelines). doi: 10.15465/gesis-sg_en_035

Behr, D., Dept, S., & Krajĉeva, E. (2018). Documenting the survey translation and monitoring process. In T. P. Johnson, B.-E. Pennell, I. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC)* (pp. 341-356). New York: John Wiley & Sons.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40,* 55–75. doi: 10.1146/annurev-soc-071913-043137

European Social Survey (2020). ESS Round 10 Translation Guidelines. London: ESS ERIC Headquarters.

Fitzgerald, R., Widdop, S., Gray, M., & Collins, D. (2011). Identifying sources of error in cross-national questionnaires: Application of an error source typology to cognitive interview data. *Journal of Official Statistics*, *27*, 569–599.

Haas, H. (2009). Übersetzungsprobleme in der interkulturellen Befragung. *interculture journal: Online-Zeitschrift für interkulturelle Studien*, *8*(10), 61-78.

Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. J. R. van de Vijver, & P. Ph. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 35–56). Hoboken, NJ: Wiley.

Harkness, J., Pennell, B.-E. & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 453–473). Hoboken, NJ: Wiley.

Harkness, J. A., Villar, A. & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 117–140). Hoboken, NJ.

Keck, V., Behr, D., & Dorer, D. (2021). *D4.9 Guidelines on the use of Translation Memories in survey translation*. doi: http://dx.doi.org/10.5281/zenodo.5180976.

Mohler, P., Dorer, B., de Jong, J., & Hu, M. (2016). Translation. *Guidelines for Best Practice in Cross-Cultural Surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from http://ccsg.isr.umich.edu/

Moorkens, J., Toral, A., Castilho, S., & Way, A. (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces, 7*(2), 240-262. doi: 10.1075/ts.18014.moo

Nitzke, J., & Hansen-Schirra, S. (2021). *A short guide to post-editing* (Vol. 16). Language Science Press. Retrieved from https://library.oapen.org/bitstream/handle/20.500.12657/52585/external_content.pdf?sequence=1

Pan, Y., & de la Puente, M. (2005). *Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials: Documentation on how the Guideline was Developed*. (Survey Methodology # 2005-06). Statistical Research Division, U.S. Bureau of the Census. https://www.census.gov/library/working-papers/2005/adrm/rsm2005-06.html

Zavala-Rojas, D., Behr, D., Dorer, B., Sorato, D., & Keck, V. (forthcoming). Using Machine Translation and Post-editing in the TRA(P)D approach: effects on the quality of translated texts. *Public Opinion Quarterly*.

Zidan, A., Awaisu, A., Hasan, S., & Kheir, N. (2016). The Living with Medicines Questionnaire: translation and cultural adaptation into the Arabic context. *Value in Health Regional Issues*, *10*, 36-40.