Questionnaire translation in cross-national survey research: on the types and value of annotations

Behr, Dorothée; Scholz, Evi

Questionnaire Translation in Cross-National Survey Research

On the Types and Value of Annotations

Fragebogenübersetzung in der international-vergleichenden Umfrageforschung

Typologisierung und Nutzen von fragespezifischen Übersetzungsanweisungen

Dorothée Behr und Evi Scholz

Abstract
Good questionnaire design and high quality questionnaire translations are vital for data comparability in cross-national survey research. With the aim to ensure comparability and prevent unintended deviations, major academically-driven studies such as the International Social Survey Programme (ISSP) or the European Social Survey (ESS) annotate the source questionnaire specifically for translation, thus providing guidance on what needs to be considered in translation. This paper studies these *translation annotations*, a topic having received scant attention in research so far. The goal of this paper is to raise awareness on this special support structure in comparative research as well as on potential pitfalls in questionnaire translation. To this end, first, translation annotations, mainly from the ESS and the ISSP, are analyzed with a view to setting up a classification of translation annotations. Second, examples of annotation types are presented together with what

Zusammenfassung
Um die Vergleichbarkeit von Daten in der international-vergleichenden Umfrageforschung zu gewährleisten, sind ein gutes Fragebogendesign beim Ausgangsfragebogen und eine hohe Qualität der Übersetzung des Fragebogens unerlässlich. Mit dem Ziel, Vergleichbarkeit zu sichern und Abweichungen zu vermeiden, kommentieren bedeutende akademische Umfrageprogramme wie das International Social Survey Programme (ISSP) oder der European Social Survey (ESS) ihre Ausgangsfragebogen speziell für die Übersetzung. Dieser Artikel untersucht diese fragespezifischen Übersetzungsanweisungen, die bisher in der Forschung kaum behandelt wurden. Ziel des Artikels ist es, diese besondere Form der Unterstützung in der vergleichenden Forschung ins Bewusstsein zu rufen sowie auf mögliche Probleme oder Fehlerquellen in der Fragebogenübersetzung hinzuweisen. Zu diesem Zweck werden erstens Übersetzungsanweisungen, hauptsächlich aus dem ISSP und

they meant for translation. Third, merits of annotations and potential criticism are discussed. Fourth, guidelines are ventured for writing translation annotations as well as for working with them. Fifth, research fields are listed in order to further explore the issue of translation annotations and its impact on translation and comparability.

dem ESS, analysiert. Hieraus resultiert eine Typologisierung von Übersetzungsanweisungen. Zweitens werden Beispiele für verschiedene Typen von Übersetzungsanweisungen zusammen mit ihrer Bedeutung für die Übersetzung präsentiert. Drittens wird der allgemeine Nutzen von Übersetzungsanweisungen, aber auch mögliche Kritik diskutiert. Viertens werden Richtlinien sowohl für das Schreiben von Übersetzungsanweisungen als auch für deren Gebrauch vorgeschlagen. Und fünftens werden Forschungsfelder aufgezeigt, die sich zur weiteren Untersuchung der Thematik der Übersetzungsanweisungen sowie zu deren Einfluss auf Übersetzung und Vergleichbarkeit anbieten.

1 Introduction

In cross-national surveys that are based on the "ask-the-same question" approach (Harkness 2003: 35), that is, where across all participating countries the same questions are asked, both a high quality source questionnaire and high quality translations are central to the overall comparability of survey data. During source text design, items should ideally be developed that match scientific needs, cross-national relevance, and translatability. To reach these goals, cross-cultural collaboration and input in formulating and selecting items for the source questionnaire is the method of choice (different forms of this type of collaboration and input are outlined in Dean et al. 2007; Fitzgerald et al. 2009; Harkness 2008a; Harkness et al. 2003; Mapi n. d.; Smith 2003). During translation then, the items should – despite the change of culture and language – maintain the intended meaning and the measurement properties of the source questionnaire. Special translation and assessment methods have been developed over the past decades in order to ensure this form of translation quality (Behr 2009; Harkness 2003; Harkness/Pennell/Schoua-Glusberg 2004; Harkness/Schoua-Glusberg 1998). With the aim to aid and "standardize" translation and to do this *before* the actual translation begins, major studies, notably the academically-driven programmes International Social Survey Programme (ISSP) and European Social Survey (ESS) or the more policy-driven Programme for International Student Assessment (PISA), have come to use *translation annotations*, that is, written aids in various forms supplementing the source ques-

tionnaire and supporting translation.[1] These translation annotations are mainly developed within the context of collaborative cross-cultural questionnaire design and draw their content from the various methods employed, such as expert review, advance translation, translatability assessment, or cross-cultural pretesting (e. g., Harkness 2008b; Jowell 1998). The overall goal of the annotations is to ensure that the translated items measure what they are supposed to measure and that equivalence to the source questionnaire is not lost because of a flawed translation.

Translation annotations, even though used in large-scale projects, have hitherto received scant attention in research. An exception being Pan, Kleiner, and Bouic (2007) and Kleiner, Pan, and Bouic (2009), respectively[2], who conducted an experimental study on the use of translation annotations in connection with more comprehensive translation instructions, and who concluded that annotations and instructions do indeed impact on translation but differently across target languages and to a yet unknown effect. In the light of these findings, Kleiner, Pan, and Bouic call for careful considerations when providing such guidance.

This article aims at raising awareness on translation annotations as a special support structure in comparative research as well as on potential pitfalls in questionnaire translation. To this end, (1) the article provides a classification of translation annotations, largely based on the ESS and the ISSP; (2) it looks into specific annotations and what they meant for ESS and ISSP translations; (3) it discusses the merits of annotations and potential criticism linked to them; (4) it ventures guidelines for questionnaire designers on how to write translation annotations and for translators on how to work with translation annotations; (5) it concludes with further research needs regarding translation annotations.

Before going into detail, a note on the terminology seems apt. What are generically called *translation annotations* in this article may have different names, and possibly different scope, in actual use. The ESS uses the term *translation annotation* (Harkness 2008b), thus lending its terminology to this paper. At the same time, the ISSP speaks of *translation or clarification notes*[3] (e. g., ISSP), and also

1 Also the Comparative Study of Electoral Systems (CSES) (http://www.cses.org/) uses so-called *notes* in their questionnaires. These to some extent cover what we focus on in this paper. They also go beyond this paper such as when, within the realm of output harmonization, concepts are defined for which countries are to produce their own wordings. In these cases we rather speak of country-specific translation of concepts into items rather than translation within the scope of the "ask-the-same question" approach. Overall, in the CSES there is a stronger emphasis on country-specific design which is expressed in the *notes*.

2 In most cases, reference is made to the 2007 report rather than the 2009 paper, since the report is more detailed and contains also appendixes with the study material.

3 These two names already suggest that the scope of notes in the ISSP may be larger than "just" pertaining to translation.

in PISA reference is made to *translation notes* (Organisation for Economic Co-operation and Development 2005, 2009). Pan et al. (2007) refer to *question-by-question explanations* (QxQs). Regardless of the name, what information do all these annotations carry?

2 A Classification of Translation Annotations

To find out what is covered by the different types of annotations, translation annotations from the major academically-driven surveys ISSP and ESS were analyzed and categorized.[4] The fact that these surveys are at the forefront of developments in comparative questionnaire design makes them a suitable ground for the analyses. In the ESS, the three core modules A, B, and C from the Round 2 questionnaire[5] (European Social Survey 2004) as well as the rotating modules D and E from the Round 4 questionnaire (European Social Survey 2008) were analyzed in terms of their content. In the ISSP, four surveys were chosen for analysis of their translation annotations. These are the 2004 survey on *Citizenship* (first module of its kind), the 2005 survey on *Work Orientation* (3rd partial replication), the 2006 survey on the *Role of Government* (4th partial replication of module), and the 2007 survey on *Leisure Time and Sports* (first module of its kind). Overall, we assume that knowledge and experiences from previous rounds or waves have found their way into the translation annotations of these recent ESS and ISSP surveys.

What is the goal of the annotations in these studies? On its cover page of the Round 2 questionnaire, the *European Social Survey* provides a definition of ESS translation annotations (2004: 2):

> "Throughout the questionnaire, annotations are provided to aid translation. These attempt to avoid ambiguity by providing definition and clarifications of the concept behind questions, especially where the words themselves are unlikely to have direct equivalents in other languages. Annotations are also used to mark new questions and questions from the Family Work and Well-being module that are located in the core."

4 The published PISA questionnaires (http://nces.ed.gov/surveys/pisa/questionnaire.asp) do not contain translation notes; therefore, PISA questionnaires could not be consulted for this paper.

5 The core modules are repeated in each ESS round. The ESS Round 2 questionnaire was chosen in view of subsequent analysis of translations of annotated items: For Round 2, four French and four German translations are available, whereas for Round 1, where the core questionnaire was set up, only three German translations are available. The Round 2 annotations are equal to the Round 1 annotations.

Learning from previous rounds, these introductory remarks have slightly changed over ESS surveys, but they still give some general information on the role of translation annotations in the ESS. To obtain a more detailed picture of translation annotations, the following classification was set up, mainly based on analysis of annotations within the above mentioned studies:

- Semantic-Pragmatic Annotations: Providing any of the following information/instructions for a *specific* term or phrase:
  - a synonym, a paraphrase, a definition, examples, implied or intended meaning components;
  - (additional) information on what is *not* included in or *not* meant by a given term or phrase;
  - explicit instruction on what should be taken into account when translating (e. g., that certain words are central in an item, that certain words should be avoided);
  - explicit instruction that countries can translate a term x by translating the alternative term y, if appropriate or more suitable in their language.
- Adaptation Annotations: Providing an instruction on permissible or required adaptations (including elements that – at the discretion of the individual countries – can either be translated or left out).
- Consistency Annotations: Providing an instruction in one of the following forms:
  - reference establishing link *within* a round or wave (e. g., to ensure consistent use of response scales within a module);
  - reference establishing link *between* rounds or waves of a survey (e. g., former rotating module question now core question; key term or phrase used already in previous rounds; item battery already used in previous rounds)[6].
- Routing Annotations: Clarifying routing (e. g., through indicating that respondents who have answered question x with 'yes' now get the following set of questions).
- Conceptual Annotations: Providing an *overall* concept definition or providing the rational for asking the question.

6 However, attention should be paid to the fact that what was working in a previous context may not work in a new context for various reasons. This especially refers to individual key terms or phrases. Embedded in a different survey question and a different questionnaire context, they may need to be translated differently compared to the previous use.

2.1 Semantic-Pragmatic Annotations

Semantic-pragmatic annotations are those that occur most often. Given the goal of questionnaire translation, namely to maintain the intended meaning of the source questionnaire, this comes as no surprise. Semantic-pragmatic annotations indicate what is meant in survey measurement terms by a given word or phrase (Harkness 2008b), what the intended and most salient reading of a question is (Harkness/Schoua-Glusberg 1998), or what the intended meaning of an item is (Harkness/Schoua-Glusberg 1998).

Work – should it be understood as paid work or any kind of work including housework and care of children? *Holidays* – should it be understood in the sense of Christmas and All Hallows or in the sense of vacation? *Concerns* – meant in the sense of worries or in a more neutral sense? Depending on what is meant translators will most often need to choose among several translation options. Semantic-pragmatic annotations can help to make the right choice, granted that a problem has been anticipated, of course. Furthermore, some semantic-pragmatic annotations may indicate the degree of freedom permitted in translation, as will be shown below.

We will now turn to different types of semantic-pragmatic annotations, first to those which offer synonyms, paraphrases, definitions, or implied or intended meaning components and which are most typical. We will thereby examine what they (can) mean for translation. The first example stems from the ESS, item C4:

> Compared to other people of your age, how often would you say you take part in social activities (27)? Please use this card.
>
> Annotation (27): Events/encounters with other people, by choice and for enjoyment rather than for reasons of work or duty.

What is the scope of "social activities"? If such a term were embedded in a normal text, its meaning could easily be identified by translators because of the surrounding context. The lack of context in questionnaires, however, poses a particular challenge for translators and so the annotation is welcome, especially given that comparable measurement depends on the right interpretation of the term. The annotation specifies the kind of activities involved, together with the contextual meaning of the adjective "social". French translations from Belgium, France, Luxembourg, and Switzerland thus render the term by "activités sociales" [*social activities*] or "vie sociale" [*social life*][7]. The German translations of Austria, Germany, Luxembourg, and Switzerland[8] use the phrase "gesellige Ereignisse oder Treffen"

7 The translated questionnaires of the ESS can be accessed at http://ess.nsd.uib.no/.
8 It is not clear, however, to what extent they collaborated and jointly produced their translations.

[*sociable events or meetings*]. The specification of events and meetings in the German wording may have been triggered by the translation annotation. Likely as it may be, however, such a link can only be hypothesized because translation process data revealing how the translation came about are not available (see for translation process research Behr 2009). The vague term "Aktivitäten" [*activities*] may have been considered to be inappropriate or misleading in the German context.

A special category of semantic-pragmatic annotations are those which, in addition to what *is* meant, make explicit what *is not* meant. By providing this comparison the actual meaning can be made even more explicit. Besides, the reason for adding the annotation can be traced back by translators, which can also aid translation and avoid misunderstanding of the annotation's content itself. The following example is taken from the ISSP 2004, item 2:

> There are different opinions as to what it takes to be a good citizen. As far as you are concerned personally on a scale of [...], how important is it:
>
> Never to try to evade taxes
>
> Annotation: **Evade** has the connotation of illegality in not paying taxes owned, and does not mean "avoid", since "tax avoidance" is not illegal.

A potential misunderstanding of "evade" was anticipated. The provision of the intended meaning helps to focus the translations on the illegal aspect.

Another type of annotation makes clear what should be taken into account when translating, e. g. in terms of terminology or register. Such annotations fulfill a highlighting function. For instance, in the ISSP 2004, item 18 asks whether the following activity was done in a given time span:

> Contacted, or attempted to contact, a politician or a civil servant to express your views.

It is supplemented by the following annotation:

> Civil servant should be translated with the appropriate term for the public service. Do not use the term "bureaucrat."

This annotation suggests that the word "civil servant" (≠ "bureaucrat") had been chosen deliberately and that this deliberate choice was felt necessary to be conveyed to translators. If during design certain word choices in the English language are clearly preferred over others, this knowledge can be communicated to translators so that, if appropriate for the target language context, they can replicate these choices. After all, the use of differently connotated words can make a difference in measurement. For example, attitudes differ depending on whether the wording "people on welfare" or "the poor" is used in an item (Weisberg 2005: 103). French- and German-language translations of the above ISSP item refer to

"Beamte" and "haut fonctionnaire", respectively. Certainly, in this and all other cases we cannot say that only because of the annotation the translation rendering is correct. Especially in the case of "social activities", however, our experience with questionnaire translation in large international studies tells us that meaning problems and potentially biased data would have arisen without the annotation.

Annotations do not only provide definitions in the form of synonyms, paraphrases or example lists; they also *explicitly* offer alternative source wording for translation. This is illustrated by the following example from the ISSP 2006, item 3b:

> There are some people whose views are considered extreme by the majority. Consider people who want to overthrow the government by revolution. Do you think such people should be allowed to ... publish books expressing their views?
>
> Annotation: In Q3b, 'publish books' can be translated as 'have their books published'.

Equally in the ISSP 2006, a scale is annotated:

> Almost none/a few/some/quite a lot/almost all
>
> Annotation: Precode: if necessary, the difference between 'a few' and 'some' can be clarified by using a term such as 'only a few'.

Scales are a particularly critical issue in translation. It is often difficult to closely match the different dimensions (agreement, satisfaction, etc.), negations (disagree, unsatisfied, etc.), or quantifications (strongly, somewhat, etc.) on the semantic level and at the same time maintain the measurement properties of the questionnaire (Behr 2009; Harkness 2003). By providing a suggestion on how to translate, the above annotation supports countries in producing a translation with appropriate distances between the scale points, especially between "a few" and "some". A very close semantic translation of the scale might end up in non-differentiation between these scale points and thus in measurement problems.[9]

Written aids *explicitly* offering alternative source formulations fulfill two functions. On the one hand, they further define meaning. On the other hand, they provide countries with a second or alternative source thereby clarifying the leeway possible in translation. This second reason has probably influenced the choice of instrument design in PISA where two source versions, one in English and one in

9 See also Fitzgerald et al. (2009) and Behr et al. (2008) on the issue of 'some' versus 'a few' – a difficult issue in scale translation!

French, have simultaneously been used since PISA 2000.[10] In this context, the PISA technical report can be quoted as follows:

- Many translation problems are due to idiosyncrasies: words, idioms, or syntactic structures in one language appear untranslatable into a target language. In many cases, the opportunity to consult the other source version may provide hints at solutions.
- The desirable or acceptable degree of translation freedom is very difficult to determine. A translation that is too faithful may appear awkward; if it is too free or too literary it is very likely to jeopardise equivalence. Having two source versions in different languages (for which the translation fidelity/freedom has been carefully calibrated and approved by consortium experts) provides national reconcilers with accurate benchmarks in this respect, and that neither back translation nor double translation from a single language could provide. (OECD 2009: 88)

Since striking a balance between faithfulness and fluency, and since identifying the needed degree of closeness (in form and content) and the possible degree of freedom (in form and content) is challenging, help in the form of approved alternative source versions in English might be useful where otherwise problems may arise. As to the issue of what is permissible in translation, Harkness/Schoua-Glusberg shall equally be quoted who already back in 1998 state that documentation could or should include, amongst others, "what is possible in terms of translation versus other forms of adaptation" (97). This statement leads us straight to the next type of annotations.

2.2 Adaptation Annotations

This group of annotations includes instructions for cultural adaptation. Intercultural questionnaire design teams should work towards preparing the source questionnaire to such an extent that anticipated permissible adaptations are marked. In this context, item 6g from ISSP 2006, received the following annotation:

> 6g refers to unemployment benefits. If there are no such benefits within a country (this applied to the Philippines in 1996), the question should not be asked.

The more cross-cultural input is received during development of the source questionnaire, the better either the source questionnaire is made to match diverse cul-

10 The principle of two source versions has in fact also been followed by the *Eurobarometer* of the EU since the 1970s.

tures or, if this does not work, the better the source questionnaire can be prepared in terms of permissible adaptations. This does not exclude the possibility that translators identify the need for further adaptations once the source text is finalized and translation is actually conducted.

The following types of annotations were partly identified not by looking at the source questionnaire documents themselves (as downloaded from the website) but by analyzing additional study material, such as specifically prepared translation templates (e. g., European Social Survey 2009) or background information on questionnaires. Equally, own translation experiences in the ESS, the ISSP, and further international studies were a helpful source.

2.3 Consistency Annotations

Consistency annotations refer to more "mechanical" aspects of the translation process. They support translators with regard to consistency *within* and *across* survey rounds. Translation annotations *within* a questionnaire can signal the repetition of scales or that of key terms. However, countries and translators must be aware of the fact that where one language can use the same term or wording throughout different linguistic contexts, other languages cannot. To take a simple example, the scale labels "good" and "bad" can be applied in the English language to all nouns irrespective of gender or number. In other languages, the adjectives need to be adapted in terms of gender and number to the corresponding noun. Or let us take a different example: A response scale may apply to a battery of 5 items. It is translated in relation to the first item and then not considered any more in the knowledge that it is the same scale for the rest of the items. The translation of the scale chosen may not fit with the other items, though. Re-translating is needed taking into account the entire item battery context; or, if needed and approved by the designers or central coordinating office of a study, a slightly different response scale is necessary for some of the items. On the whole, it should be well considered where, if at all, consistency notes *within* a questionnaire are offered. If offered, they should in no case lead to false security and thoughtless copy-paste activities among translators.

Consistency aids are in fact a requirement when it comes to replicating entire modules or parts of these *across* survey rounds or waves. The goal to measure change over time makes it necessary that translations are re-used unless serious mistakes are found. Otherwise it is not clear whether changes in results are due to changes in the stimulus or due to real changes in attitudes, behavior, etc. Besides, countries need to be provided with clear instructions on whether a trans-

lation is to be re-used unchanged (in the case of a real item replication) or whether it needs slight modifications (if the source item itself has undergone slight modifications). There are different ways available as to how replication across rounds or waves can be signaled. Separate documentation, prefixes or similar methods are possible.[11]

2.4 Routing Annotations

In questionnaires where rather complex routing is used (e. g., due to the work situation of respondents), references to preceding questions may ensure understanding in context and thus facilitate translation. It is easy to imagine different scenarios, in which routing takes place, such as: 'questions 1-4 are addressed to self-employed respondents whereas questions 5-12 are addressed to employed respondents' or 'question 10-12 are asked when question 1 is answered with 'no''. Knowing about this will help translators to maintain consistency in terminology and to ensure the questionnaire flow for respondents.

2.5 Conceptional Annotations

Conceptional annotations, as understood in this paper, refer to the overall theoretical concept or dimension to be measured, or the rational for asking specific questions. With rare exceptions, this information is not included in the source questionnaire documents themselves which were looked at for this analysis. Conceptional information may overwhelm translators, especially if they are not knowledgeable of social research. It shall be argued here that it makes sense to provide this information, if not to translators, than at least to cross-national research teams who are responsible for overseeing and signing off the translations, and who are most likely to be social scientists.

First, conceptual information enables cross-national teams to assess whether a given question can be translated following the general or current understanding of an *ask-the-same question* approach (see e. g. Behr 2009 or Harkness et al. 2004 for more specific information) or whether the question needs to be adapted and rewritten in order to measure the defined concept in the target culture. Harkness (2008a: 74) provides the following example: "Do you have difficulty reading a newspaper, even with spectacles?" If newspapers or vision aids are not readily available in certain cultures, the most salient reading (in a close

11 E. g. from the ISSP 2006: http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/issp/excerpts/excerpt_gmr_2006.pdf.

translation) will probably diverge from the intended reading: rather than collecting information on whether or not the respondent's vision is impaired, his or her reading skills may be assessed. In such cases, knowledge of the underlying concept helps to identify the need for reframing and adapting a question to make it suit the target culture. It needs to be said, however, that if a survey is designed as a comparative survey right from the start and if *asking-the-same-question* is the preferred approach in producing the measurement instruments in all countries and cultures, issues such as the one above by and large should ideally have been identified during design and either improved for cross-cultural application or earmarked for country-specific adaptation.

Second, conceptual knowledge may lead to more effective review processes. Behr (2009) shows that quite some time may be spent in team review processes on questioning the concept behind a given item. These discussions are certainly more likely if subject matter or design experts are involved in review processes, that is, those knowledgeable of the relationship between concepts and questions. A concept description, if available, can shorten conceptual discussions during review processes. However, a note of caution is also apt: Having a concept description at one's disposal can equally lead to lengthy discussions since review participants might not agree with the operationalization from concept into questions (notwithstanding cross-cultural concerns). If the source text is already finalized and cannot undergo change anymore, these discussions mostly cannot result in any changes, though. Thorough documentation of concerns and feedback to the questionnaire designers should then be the preferred approach to give designers and analysts the chance to learn from this.

While the above sections focused on a classification of annotations, the following section discusses more in general the value of providing annotations but also addresses potential criticism related to annotations.

3 Discussion

Critical voices may reject the use of annotations, particularly that of semantic-pragmatic annotations. They may say that the source question should be clear enough in the first place because otherwise English-language respondents will not know either what is meant, and so validity is put at risk in any case. Such a statement can fully be endorsed, but there are also qualifications to be made.

First, the source questions may be as well designed as possible in the sense of reducing ambiguity of meaning. However, different readings can always occur

because meaning is not static but determined by context, personal characteristics, etc. (Harkness/Schoua-Glusberg 1998). To assure that indeed the intended and most salient meaning is conveyed in the translation, an annotation may help.

Second, despite good foreign language skills, translators may not always perceive what English native speakers perceive as the most salient reading of an item (Harkness et al. 2004). Using the scenes-and-frames-semantics by Fillmore as an explanatory framework for this (e. g., Kußmaul 1995, 2007), one can also say that the scenes activated in the minds of non-English native speakers when reading or hearing a given item may not be equivalent to the scenes activated by English native speakers. The better the linguistic and cultural knowledge of translators, the closer their understanding is to native-speaker understanding or the more at least they are in a position to identify interpretation problems. In order to avoid any form of misunderstanding and, consequently, loss of item comparability, a translation annotation can be provided.

Third, it seems that often people inexperienced in translation translate source questionnaires into the target language (e. g., Hambleton 2005; Harkness et al. 2004; see also Jowell 1998). There is evidence available in translation research (e. g., Krings 1986; Kußmaul 1995) that inexperienced translators often follow too closely source text structures or tend to provide only fixed one-to-one equivalencies (those that they have learned as direct equivalences in school, for instance) that may not be appropriate in the given context. For those people, translation annotations are especially helpful since they prevent inexperienced translators from choosing potentially wrong automatic word-for-word replacements. For instance, if an item reads "failing to declare your income", then this "failing" should not be understood as "trying but not succeeding", which might be the most prominent reading of "fail" for inexperienced translators who are not native speakers of English. It should rather be understood as "not doing what is expected or needed". An annotation can clarify this meaning and thus prevent misunderstanding. However, needless to say translation annotations in general should not become the means to make up for incompetent translators. Highly skilled practitioners should be the rule in survey translation in order to ensure translation quality, and ultimately, quality of substantive conclusions drawn from cross-cultural survey data.

We would now like to address potential criticism of translation annotations or potential limitations. First, cases have occurred where the annotations have led to quite explicit translations through full integration of the annotation into the item – a fact which has lead to concerns over translation annotations (see also Pan et al. 2007 on this issue). The explicit integration can be illustrated with the following example on religious identification that is measured in the ESS with item C9:

> Do you consider yourself as belonging to (32) any particular religion or denomination?
>
> Annotation (32): Identification is meant, not official membership.

The German and Austrian team both integrated the annotation's content into the item by asking:

> Unabhängig davon, ob Sie Mitglied oder Angehöriger einer Kirche oder Religionsgemeinschaft sind, fühlen Sie sich einer bestimmten Religion oder Konfession zugehörig? [*Regardless of whether or not you are member (two words in German for "member") of a church or denomination, do you consider yourself as belonging to a certain religion or denomination?*]

Integrating the annotation into the item can make the intended meaning of this item clearer and may thus be truly respondent- and researcher-friendly. At the same time, this example raises the question of whether this happens at the expense of reducing measurement error in the German context compared to measurement error in the UK context. Was a close translation along the lines of the English source text considered to be confusing or unclear in German? Did the translators decide on purpose to make the non-membership aspect explicit? The Swiss-German and Luxemburg-German versions do not add the information on membership and neither do the different French versions, but this finding shall not lead to automatically rejecting one translation and favoring one over the other simply because of different levels of closeness to the source text. It may be that the more explicit German translation is indeed closer to the source text in terms of respondent's understanding than are the other translations. Effects of the difference between the various German versions and between these German versions and the English version can only be assessed empirically, though. Documentation on the difference between source text and translation would in any case have helped to understand the choices made.

The ESS is not alone when it comes to different uses of translations notes. In ISSP 2007, item 1f reads:

> Get together with relatives
>
> Annotation: who do not live in your household

Across the German-language versions[12], the translations chosen for "get together" ("-besuche", "treffen") imply that the meetings occur with relatives that one does not see regularly at home; hence, an explicit additional sentence is unnecessary.

12 Austria (CAPI), Germany (CASI), Switzerland (CAPI).

The situation is different with the French-language versions. The translators in the two French-language translations from France and Switzerland[13] opted for translating the annotation itself and adding it in parentheses to the item. For instance, the French translation from Switzerland reads "Voir des membres de votre famille (qui ne vivent pas avec vous)" [*Seeing members of your family (who do not live with you)*]. Here, the additional parentheses seem almost imperative because seeing members of your family may include daily interaction with parents or siblings. In the South African English version, the annotation was not added in parentheses but fully integrated into the item: "Get together with relatives who do not live in your household." This addition might even be required within the cultural context of South Africa. Regarding the addition of parentheses or of further information, the presented country versions differ. This might give rise to concern, at least if one expects a translation that closely mirrors the source text. However, the different country solutions may indeed cover the intended meaning in a suitable way in all the countries. What would have been the translation outcome if no clarification had been given? Would some countries have produced a translation where regular seeing of one's family at home had been (implicitly) included? This would certainly not have been the goal of the item.

In sum, quite some changes or (formal) differences compared to the source text item can happen when the content from annotations is taken into account in translation. These changes or (formal) differences should not automatically be rejected, though. They can very well be justified and identified as the best possible solution to cover the intended meaning. This being said, there may certainly be some mishandling or misunderstanding on the use of annotations, too. Instructions on how to use translation annotations when translating are thus vital to the concept of annotating the source questionnaire for translation. More on such instructions can be found further below.

Second, annotations do not automatically lead to translations conveying the intended meaning of an item. They should thus not be misunderstood as a cure-all in view of comparability. The following example from the ISSP 2004, item 4, can be used as a case in point. The item reads:

> To keep watch on the actions of government.
>
> Annotation: Keep watch means exercise vigilance in observing government, with a view to pointing out unwarranted actions or ensuring that proper actions are conducted.

13 France (mail), Switzerland (CAPI).

The idiomatic expression "keep watch" in relation to government actions is difficult to render in many languages, especially in terms of direct or formal equivalence. The vigilance aspect seems to have been expressed in the German-language translations with an adverb, for instance: "sehr aufmerksam verfolgen", "immer genau verfolgen" [*to (always) follow attentively/meticulously what the government does*]. Does this really tap the key ideas of "keep watch", though? One of the French-language translations reads: "Se tenir au courant des actions du gouvernement" [*keep up to date with what the government does*]. Although there is certainly a large meaning overlap, the vigilance aspect has probably not sufficiently been represented in the French translation, and possibly also not in the German translations. Something different may be measured by the translated items than intended. Could it be that the idiomatic expression is on the whole quite difficult to render in other languages, especially if one aims at a translation that is as concise as possible? A change in the source text rather than the annotation might have been the better choice to assure equivalence in meaning. This also brings us to the next point.

Third, annotations might be misleading if they contain information or clarification that is not – directly or indirectly – conveyed by the source text item itself. By taking into account the annotation content, the translation might well cover the concept but it might not be comparable to the source item any more. Any improvements should therefore be done on the source items themselves.

In the following two sections, we want to venture some suggestions for both writing translation annotations and using them. Hopefully these suggestions will encourage researchers to further explore the issue of translation guidance in comparative survey research. It is hoped as well that they can foster further discussion about best practices in comparative questionnaire design and questionnaire translation. As Kleiner et al. (2009) have already shown – there is much in survey translation and guidance that deserves further attention.

4 On Writing Translation Annotations

One should consider in the first place which help to provide to translators, e. g., if information on theoretical concepts is to be provided; if primarily aids clarifying meaning of individual words or phrases should be given; if consistency guidance is offered; or if routing help is needed. One may also want to point to overall design and layout issues, although the target group should be questionnaire designers then rather than actual translators in the latter cases. Overall, a proper balance needs to be found between useful information and information overflow. In this

context also comments on translation notes from evaluators participating in the study of Pan et al. (2007) are revealing. Questioned about their beliefs regarding question-by-question explanations to translators, a translation evaluator considered them "to be very helpful to translators, but they should be provided selectively and not for every question" (19).

At what stage in the process should the annotations be drafted? This depends on what is desired. Guiding information pertaining to theoretical concepts should be assembled within the process of good questionnaire design in general. It does not, therefore, require any drafting specifically for translation. The same generally also applies to routing.

Guiding information related to consistency on the "macro-level" (that is, consistency *between* survey waves or rounds) is a general issue of questionnaire design, too, because here decisions are made as to which items to replicate, which items to modify, etc.[14] Therefore, this information will also already be available.

Guiding information related to consistency on the "micro-level" (that is, consistency *within* a survey such as repetition of response scales or important reoccurring terminology within a questionnaire) is more translation-specific. It may be added to the questionnaire at the latest *after* the questionnaire itself has been drafted and *before* it is sent out for translation. Alternatively, this guidance can be produced simultaneously with the design process. Care should be taken, though, that the chosen level of consistency notes, if they are used at all, is appropriate, otherwise translation when following these notes might end up in piecemeal or bad translation. Good translators will certainly look out for consistency with or without annotations, especially if the importance of consistency, such as consistent translation of "agree-disagree" scales for measurement purposes, has been pointed out to them.

Annotations pertaining to the meaning of items, to word choices, or to adaptations should ideally be prepared simultaneously with the source questionnaire. The more intensive cross-cultural input there is during the design phase (international design and expert review teams, advance translations, etc.), the better can translation annotations in terms of meaning, word choice, or adaptations be taken into account. In the same context, knowledge of translation problems in similar or preceding studies will be helpful in annotating the source text, especially

14 The use of computer-aided translation software, including translation memories and terminology databases, is not taken into account in these descriptions on consistency. This software can support consistency within and across survey waves as well as standardization of translation and facilitates or even makes unnecessary some of the more manual checks of translation consistency. See also Bowker (2002).

when it is known that certain terms again and again pose problems in cross-cultural implementation. When clarifying meaning, it surely cannot be the goal to rephrase the entire questionnaire for people who do not properly speak the source language or have never attempted a translation. In this context it should rather be stressed that researchers or translation commissioners need to select translators with high proficiency language and translation skills and a good expertise in translating questionnaires and familiarity with social research, or to assemble translation teams that, on the whole, meet the needed expertise (see Harkness 2003 on team approaches in translation).

All in all, a good overview of translation processes and common translation mistakes is needed in order to make appropriate decisions as to when annotations are useful. Each survey has its own needs (due to different countries participating, due to the topic, the translation actors, etc.) and so the needed degree of guidance will necessarily differ between surveys.

Eventually, one should also consider in which form information should be provided. Any information offered should be both easily accessible and assembled in the least documents as possible in order to allow easy overview of translation documentation.

5 On Instructing Translators

Since translation annotations are a special text feature of questionnaires, instructions to translators on how to work with these are needed. Pan et al.'s (2007) study shows that, in addition to the actual translation annotations, instructions have an effect, so one should carefully outline how one envisages the use of annotations. Instructions handed out to translators in experimental groups in Pan et al. (2007: 6-7) are:

> "We ask that you translate the survey items with respect to their **intended meaning**. Before attempting to translate each item, read carefully the explanation that follows the item in order to get a better sense of what is intended. (The explanations are all in italics – do not translate these.)"

and

> "There are culturally different ways of using language in interaction. Please translate the survey items so that they sound as natural as possible in the context of a telephone interview. This means that the questions should not sound awkward to the survey respondents, and the questions should be phrased in a culturally appropriate way. Feel free to make whatever changes necessary to accomplish this."

In our view, these instructions do not fully tell translators what they should do with the annotations other than not translating them. For the experimental group that gets both of the instructions above (other experimental groups get different sets or combinations of instructions), the second instruction may even be read as the approval for all types of changes in translation ("Feel free to make ..."). Here, of course, one must not wonder if the translation almost becomes a text on its own with few links to the source item left. Problems with the latter instruction are in fact also discussed by Kleiner et al. (2009).

It seems pressing to address the following questions when providing translators with instructions on how to use translation annotations:

a) Are translation annotations meant as translation aid only or should the respective information be added to the questionnaire as interviewer or respondent notes, if necessary?

b) Should they "merely" help to trigger the correct translation or should they themselves in terms of their entire content, that is, literally, be integrated into the item itself, if necessary?

These questions are not trivial considering that in both the ESS and ISSP countries have dealt differently with translation annotations. These different dealings can be due to different understandings of translation annotations, to different conceptions of what a good and comparable survey translation is, to different approaches from question designers as to what annotations involve and how efficient they are, but also to the different linguistic and cultural systems involved. What works in some languages (e. g., a close translation of the source) does not work in other languages. To convey the intended meaning, some languages need to produce more elaborate wordings than other languages, thereby even taking up examples from the translation annotation. An instruction compendium that applies to all types of annotations, items, or language combinations, therefore, cannot be given. Some approximations shall be ventured, however:

Semantic-Pragmatic Annotations: In an *ask-the-same-question* approach where the same question is asked in several countries in different languages, a principal message for the use of translation annotations, following current expectations, could be to mirror the intended meaning and to do this without spelling out all the examples or meaning features that are given in an annotation. The annotation for the question

> "During the past 12 months, did (CHILD) receive all the routine preventive care that (he/she) needed?" may read: "*Preventive care" is defined as measures taken in advance by health care providers that emphasize prevention, early detection, and early treatment of illness, injury, or long term health problems."* (Pan et al. 2007, B-5)

How should translators deal with such an annotation? We would suggest that if there is a word or phrase in the target language that is usually used to denote this type of preventive care, this word should be used in translation. If, however, such a word does not exist (which may even indicate that the concept of preventive care does not exist!) or if it is not understandable by the target respondents, a paraphrase taking up some of the information from the annotation might be the only option. Still, the task should be to be as comparable *and* concise as possible *and* to help the respondent to understand and answer the question with the minimum of effort. Basic guidance such as this should be part of the general instructions to translators, though; it does not need to be included in an annotation itself. On the whole, insights into comparability requirements within cross-national survey research and into questionnaire design are deemed necessary in order to translate appropriately – and to assess annotations and their rational adequately.

Consistency Annotations: Consistency annotations can facilitate translation work. Translators should still assess whether the micro-consistency level (e. g., reoccurring terminology) from the source text, as indicated in annotations, can be maintained in the target text or whether target language grammar and idiom make a deviation unavoidable. Annotations signaling the replication of previously used translations should not prevent researchers from regularly checking that the translation wording is still up to date and in line with current developments in society.

Routing Annotations: These annotations are particularly helpful when routing is complex since they allow translation in context where, on the paper surface, the normal linear flow has been interrupted.

Conceptional Annotations: These highlight the theoretical framework behind questions. Especially for those familiar with social research this information helps to understand the purpose of the items. If, for cultural reasons, the relationship between the item (in close translation) and the concept is questioned, this should be raised with the designers, ideally before the source questionnaire is finalized. If (professional) translators are to use these annotations, they will need training on how to relate them to the items and their translations.

Despite all the value of annotations, we wish to stress that translation annotations on the whole cannot act as a substitute for subject matter knowledge, knowledge of comparability requirements, and expertise in questionnaire design and translation. This knowledge and expertise is needed in the end in order to draw the line between closeness in translation and allowed freedom in translation. And this line will always depend on which language pairs are involved in translation.

6 Conclusions

Current practice in methodological forerunner studies suggest that annotations – or whatever their name – are on the whole helpful and comparability-enhancing. Otherwise, these written aids would not still be in use in these surveys. However, some annotations are surely better than others, and this also applies to translations which draw on the help of these annotations.

Annotations mainly arise from cross-cultural collaboration. On the one hand, this collaboration makes it possible to phrase and select items that are most suitable for international comparison. On the other hand, this collaboration permits identification of issues that, despite a well-designed source questionnaire, may be problematic in translation or could potentially be misunderstood. These issues can then be "translated" into translation annotations. The overall goal of these annotations is to ensure equivalence and to ensure that data is not excluded from international comparison only because of flawed or biased translations. In any case, the better cross-cultural collaboration is integrated into the design process, the better translation and comparability needs can be catered for.

When designing translation annotations, a balance needs to be found between offering too little guidance and too much guidance. The topic, the languages and cultures involved in a survey, as well as the actual questionnaire will determine in each single case which guidance is needed. There is no universal recommendation to be made. Despite this, however, we hope that we have succeeded in raising awareness on what can be done in comparative survey research to enhance comparability and where potential pitfalls lay in translation (that hopefully can be circumvented by annotations).

To empirically determine the impact of annotations on comparability and respondent comprehension, further studies are needed where translation is done with and without these notes, where it is done by experts from different fields with or without these notes, and where subsequent assessment and cross-cultural cognitive pretesting of translated versions rounds up these studies. As Kleiner et al. (2009) rightly point out, the ultimate test on how comparable and good translations are is not a purely linguistic test based on formal differences between source questionnaires and their translations, but it rests with actual respondents. In addition, it may be worthwhile investigating systematically the impact on how much annotations lead to a greater explicitness in translation compared to the source text. Research in translation studies of translations being more explicit than the source text could be explored for these purposes (e. g., Englund Dimitrova 2005; Klaudy 1998). Knowledge of the challenging task of translation can greatly be enhanced by these studies.

References

Behr, D., 2009: Translationswissenschaft und international vergleichende Umfrageforschung: Qualitätssicherung bei Fragebogenübersetzungen als Gegenstand einer Prozessanalyse. Bonn: GESIS.

Behr, D., J. Harkness, R. Fitzgerald, and S. Widdop, 2008: ESS Round 4 Questionnaire and Translation Queries. London: European Social Survey, Centre for Comparative Social Surveys, City University.

Bowker, L., 2002: Computer-Aided Translation Technology: A Practical Introduction. Ottawa: University of Ottawa Press.

Dean, E., R. Caspar, G. McAvinchey, L. Reed, and R. Quiroz, 2007: Developing A Low-Cost Technique For Parallel Cross-Cultural Instrument Development: The Question Appraisal System (QAS-04). International Journal of Social Research Methodology 10 (3): 227-241.

Englund Dimitrova, B., 2005: Expertise and Explicitation in the Translation Process. Amsterdam: John Benjamins.

European Social Survey, 2004: Source Questionnaire, Round 2, 2004/5 Final Version, Amendment 03, 21.07.2004. London: Centre for Comparative Social Surveys. Retrieved December, 07, 2009, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=63&Itemid=356.

European Social Survey, 2008: Source Questionnaire Amendment 03 (Round 4, 2008/9). London: Centre for Comparative Social Surveys. Retrieved December, 07, 2009, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=63&Itemid=356.

European Social Survey, 2009: ESS R4 Translation Templates. London: Centre for Comparative Social Surveys. Retrieved August, 23, 2011, from http://www.europeansocialsurvey.org/index.php?option=com_docman&task=cat_view&gid=92&Itemid=80.

Fitzgerald, R., S. Widdop, M. Gray, and D. Collins, 2009: Testing for Equivalence Using Cross-National Cognitive Interviewing. City University London, Centre for Comparative Social Surveys. Working Paper Series 1. Retrieved June, 04, 2009, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=148&Itemid=340.

Hambleton, R. K., 2005: Issues, Designs, and Technical Guidelines for Adapting Tests into Multiple Languages and Cultures. Pp. 3-38 in: R. K. Hambleton, P. F. Merenda, and C. D. Spielberger (Eds.): Adapting Educational and Psychological Tests for Cross-Cultural Assessment. Mahwah, NJ: Lawrence Erlbaum Associates.

Harkness, J., 2003: Questionnaire Translation. Pp. 35-56 in: J. Harkness, F. J. R. van de Vijver, and P. Ph. Mohler (Eds.): Cross-Cultural Survey Methods. Hoboken, NJ: Wiley.

Harkness, J. A., 2008a: Comparative Survey Research: Goal and Challenges. Pp. 56-77 in: E. de Leeuw, J. J. Hox, and D. A. Dillman (Eds.): International Handbook of Survey Methodology. New York: Lawrence Erlbaum Associates.

Harkness, J. A., 2008b: Round 4 ESS Translation Strategies and Procedures. Retrieved September, 12, 2008, from http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=66&Itemid=112.

Harkness, J., B.-E. Pennell, and A. Schoua-Glusberg, 2004: Survey Questionnaire Translation and Assessment. Pp. 453-473 in: S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, and E. Singer (Eds.): Methods for Testing and Evaluating Survey Questionnaires. Hoboken, NJ: Wiley.

Harkness, J. A. and A. Schoua-Glusberg, 1998: Questionnaires in Translation. Pp. 87-127 in: J. Harkness (Ed.): ZUMA-Nachrichten Spezial 3: Cross-Cultural Survey Equivalence. Mannheim: ZUMA. Retrieved September, 19, 2011, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial3.pdf.

Harkness, J., F. J. R van de Vijver, and T. P. Johnson, 2003: Questionnaire Design in Comparative Research. Pp. 19-34 in: J. Harkness, F. J. R. van de Vijver, and P. Ph. Mohler (Eds.): Cross-Cross-Cultural Survey Methods. Hoboken, NJ: Wiley.

ISSP Research Group: International Social Survey Programme 2004: Citizenship (ISSP 2004). GESIS Data Archive, Cologne, Germany, ZA3950_bq Basic (Source) Questionnaire; doi:10.4232/1.10078.
ISSP Research Group: International Social Survey Programme 2005: Work Orientation III (ISSP 2005). GESIS Data Archive, Cologne, Germany, ZA4350_bq Basic (Source) Questionnaire; doi:10.4232/1.4350.
ISSP Research Group: International Social Survey Programme 2006: Role of Government IV (ISSP 2006). GESIS Data Archive, Cologne, Germany, ZA4700_bq Basic (Source) Questionnaire; doi:10.4232/1.4700.
ISSP Research Group: International Social Survey Programme 2007: Leisure Time and Sports (ISSP 2007). GESIS Data Archive, Cologne, Germany, ZA4850_bq Basic (Source) Questionnaire; doi:10.4232/1.10079.
Jowell, R., 1998: How Comparative is Comparative Research? American Behavioural Scientist 42 (2): 169-177.
Klaudy, K., 1998: Explicitation. Pp. 80-84 in: M. Baker and K. Malmkjær (Eds.): Routledge Encyclopedia of Translation Studies. London: Routledge.
Kleiner, B., Y. Pan, and J. Bouic, 2009: The Impact of Instructions on Survey Translation: An Experimental Study. Survey Research Methods 3 (3): 113-122.
Krings, H. P., 1986: Was in den Köpfen von Übersetzern vorgeht: eine empirische Untersuchung zur Struktur des Übersetzungsprozesses an fortgeschrittenen Französischlernern. Tübingen: Narr.
Kußmaul, P., 1995: Training the Translator. Amsterdam: John Benjamins.
Kußmaul, P., 2007: Verstehen und Übersetzen: ein Lehr- und Arbeitsbuch. Tübingen: Narr.
Mapi Research Institute (n. d.). Translatability Assessment℠. Retrieved August, 2, 2009, from http://www.mapi-institute.com/linguistic-validation/services/translatabilityassessment.
Organisation for Economic Co-operation and Development. 2005: PISA 2003 Technical Report. Paris: OECD.
Organisation for Economic Co-operation and Development. 2009: PISA 2006 Technical Report. Paris: OECD.
Pan, Y., B. Kleiner, and J. Bouic, 2007: The Impact of Instructions on Survey Translation: An Experimental Study. Survey Methodology 18. Retrieved December, 01, 2009, http://www.census.gov/srd/papers/pdf/rsm2007-18.pdf.
Smith, T. W., 2003: Developing Comparable Questions in Cross-National Surveys. Pp. 69-91 in: J. Harkness, F. J. R. van de Vijver, and P. Ph. Mohler (Eds.): Cross-Cultural Survey Methods. Hoboken, NJ: Wiley.
Weisberg, H. F., 2005: The Total Survey Error Approach: A Guide to the New Science of Survey Research. Chicago: University of Chicago Press.

Addresses of the Authors

Dr. Dorothée Behr
GESIS – Leibniz-Institut für Sozialwissenschaften
B 2, 1
68159 Mannheim
Dorothee.Behr@gesis.org

Dr. Evi Scholz
GESIS – Leibniz-Institut für Sozialwissenschaften
B 2, 1
68159 Mannheim
Evi.Scholz@gesis.org