

Lost in translation: Data quality and interpreters in multilingual research
[Towards an Interpreting Methodology]

Lauren Doerr

NORC at the University of Chicago
55 E. Monroe
Suite 1840
Chicago, IL 60603
Doerr-Lauren@norc.uchicago.edu

AAPOR Annual Conference, 2005
Miami, FL

Lost in translation: Data quality and interpreters in multilingual research

A brief glossary of technical terms: Though commonly used interchangeably, the term translation refers to the recoding of *written* text from one language to another; interpretation is the transformation of *verbal* speech.

With the surge of limited English speaking populations in the U.S., as well as the growth of multinational studies, the need for multilingual surveys is established and growing. The administration of surveys in multiple languages has opened up a brave new world of methodological dilemmas, questions and potholes as well as benefits. Recent work in survey research, health research, and other disciplines has done much to establish a rigorous and valid methodology for questionnaire translation, but little has addressed the growing phenomenon of telephone interpreter use in multilingual surveys. I will suggest that virtually all of the issues and pitfalls raised in translation methodology literature also apply to interpretation, but with fewer opportunities to intervene and analyze. However, there are steps that the researcher can take to maximize consistency of administration with telephone interpreters, and hence higher data quality.

The growth of multilingual studies both domestically and internationally has led to many recent efforts to standardize translation procedures, and develop an effective methodology for translating questionnaires in survey research, health research, and other areas. These methodologies are unified by similar elements which at minimum include:

- Teamwork in translation (either instruments are independently translated, then reconciled, split translation, or translated in a team setting)
- Review of problem areas in translation
- Revision of translation incorporating changes made during review process

When executed with proper translation methodology, the translated questionnaire is expected to be “equivalent”, and thus functions identically, or as closely as possible, as the original. Questionnaire “equivalence” suggests that the translated question: says the same thing, means the same thing, measures the same thing, measures equivalently, imposes the same burden on respondents, and meets reliability and validity requirements (Shoua-Glusberg 2004). Respondent answers from any language group, can then be analyzed individually, or pooled with other populations. Rigorous methodology helps to ensure that comparisons made between linguistic and cultural subgroups are accurate.

Though no methodology has been widely established for the use of verbal interpreters, they can also be of use to the survey researcher. If a project does not know what languages it will encounter in the field, wants to conduct only a few interviews over a great multitude of languages, or has any other reason that makes a full instrument translation unfeasible, interpreters can greatly expand the reach of a survey to limited English speaking respondents. Telephone interpreters in particular can be an effective way to reach many more individuals. Companies such as AT&T Language Line, Bowne

Global Solutions, Tele-Interpreters, and others offering “on-demand” telephonic interpretation are growing and multiplying and can be a major aid to a survey operation.

The use of telephone interpreting services in research is growing, perhaps because of their ease of access, low initial cost outlay, and immediacy. Beginning at AAPOR in 1995, telephone interpreters have been presented as an integral part of certain surveys, especially RDD studies. Without prior knowledge of what languages field staff will encounter, telephone interpreters allow an interviewer the flexibility to field virtually any language, with little initial cost outlay. With one phone call, the interviewer can go from hearing gibberish on the other end, to getting help in identifying the language a potential respondent speaks, communicate with an individual over the phone, and complete an interview.

There is virtually no literature on interpreter usage in surveys and its impact on data quality, but many lessons from translation methodology can be applied to interpreter use: the usefulness of developing questionnaires with wording that will facilitate translation as well as cultural adaptation (Harkness, van de Vijver & Johnson, 2003), the challenges of “asking the same question”, the necessity of careful evaluation of response bias in certain cultures (Smith, 2003, 2004), and hazards of poorly executed work. As with translation, it is important to use qualified professionals rather than any bilingual individual to interpret.

But interpretation can be an unruly, unpredictable cousin to translation, with numerous additional methodological challenges. In translation, once the instrument is rendered in the target language, the interview proceeds similarly as would an English interview. The respondent and interviewer speak the same language, and the interviewer reads the questions, as well as records respondent answers. The interviewer is the individual probing, or coding the response into an answer choice. Interpretation differs in that respondent and the interviewer do not speak the same language, so the interview becomes triangulated. All verbal exchanges, the questionnaire and probes as well as respondent answers must first pass through the filter of the interpreter.

The elusive nature of interpreted speech

Recall the efforts made in survey research, health studies and other disciplines to standardize translation procedures. Each important procedural step hinges on the fact that the translation is written. The same version can be read by all reviewers, it can be discussed and then altered, and a defined result is what reaches the field. Then, since all speakers of one language receive the same translated questionnaire, any errors not caught during the review process may reveal consistently questionable data. Not so with verbal interpretation. In contrast to translated questionnaires, which can be administered consistently like their English counterparts, each interpretation will vary slightly. Even the most professional of interpreters may interpret the same question using slightly different wording in different interviews.

What happens in interpretation when items are presented that even a translation team, with the ability to review and evaluate, find challenging? Take, for example, some items

in a health questionnaire translated in writing by FACIT Translation at Northwestern University. These items were translated from US English, then tested in cognitive interviews before being reviewed again and revised by a team of translators:

- US English: “How often did you have trouble walking a block or climbing one flight of stairs?”
- German: “Wie oft hatten Sie Schwierigkeiten, einen Häuserblock (100 Meter) weit zu gehen oder eine Treppe hochzusteigen?”

During testing, German patients often interpreted this to mean “around the block,” which is equivalent to about 400 meters.

- US English: “I feel dull and sluggish.”
- Finnish: “Tunnen itseni tylsäksi ja saamattomaksi.”

During testing, Finnish patients interpreted the original translation of “dull” to mean “socially uninteresting”, and the translation was revised.

Similarly, American colloquialisms such as “I feel full of pep” or “feeling blue” in the SF-36 Health Survey have been shown to be problematic in translation, but when revised as “full of life” or “depressed”, can be translated with equal validity to the English originals (Bullinger et al, 1998). How would these questions come across during an interpreted interview? In interpretation, both the instrument and the responses are being transformed on the spot by the interpreter, and there is no record of the words that the respondent heard or said.

The transient nature of speech also means that there is no way to review if there are any strange data patterns, and errors may not be universal throughout a language. Harkness, Pennell and Schoua-Glusberg (2004) cite an example where an error in translation rendered a German version of a question that was substantively different than the English original. Respondents were asked about their views on “organizing protest marches that obstruct/block traffic” rather than the original “organizing protest marches and demonstrations.” The error was only caught after noting that data patterns for German and Austrian respondents differed substantially from those of other countries. Careful examination of the translated questionnaire yielded the discrepancy. In interpretation, this would be an untraceable error that also might not have occurred with every German respondent, further muddying data.

Survey interpreting is unique

The demands of survey interpreting are quite different than other types of interpreting. “Good interpreters” from other fields may actually hinder the work of the survey researcher. Medical interpreters or customer service interpreters are primarily concerned with the transmission of meaning, not the precision of the interpretation. A “good interpreter” for the anthropologist or nurse is one who can bridge the cultural gap by ironing out communication difficulties. With communication as her goal, a “good interpreter” may explain complicated or technically worded survey questions in a simpler way to a lesser-educated limited English speaking respondent. If the parallel English speakers, or more highly educated language speakers do not receive the same

explanations, this may bias survey results, or bias only the results of those who have worked with that interpreter. A “good interpreter” may eliminate the leading phrase “some people do not take their medications as they are supposed to” if she feels that will offend the respondent, or might only translate the applicable word in “do you attend services at a church, temple, mosque, synagogue, shrine, or other religious center?”, violating survey protocol.

As has been explored in the translation methodology literature, a translator is not versed in the “rules” of survey discourse. He or she also makes decisions about how to translate based on her perception of the intended meaning of the question. In translation, these issues can be addressed on a question by question basis, either through notes to the translator, or by committee review. An essential step in good translation methodology is the ability to give translators specifications, the translators’ ability to ask questions as they work, and the ability to review the result. An interpreter can often be on her own.

Words, words, words...the precision of survey language

We know as survey researchers that questionnaires are not worded for the transmission of meaning only; rather, they are carefully constructed and deliberate in their word choices and order. This is for good reason; experimental studies have confirmed the effect of questionnaire wording on response patterns.

Smith (1987) and Rasinski (1989) demonstrated that two ways of wording similar concepts could trigger differing reactions from respondents. In an experiment conducted on the General Social Survey (GSS), respondents were asked one version of the questionnaire that referred to their support for “welfare”, and in another version, were asked about support for “caring for the poor” and “assistance for the poor”. Respondents were more supportive of “assistance for the poor” than for “welfare” (Smith 1987). Similarly, Rasinski (1989) demonstrated that “halting rising crime rate” and “dealing with drug addiction” generated more support than “law enforcement” and “drug rehabilitation”, respectively.

Even minor wording changes have been shown to affect response. In exploring small modifications meant to better tailor a questionnaire to its respondents, Rasinski, Ernst and Haggerty (2003) demonstrated that specifying actions or adding clauses could compromise the comparability of a question item to the original data set. “I live in a close-knit neighborhood.” was more easily agreed to by respondents in relation to the other items on a scale than was the sentence “this is a close-knit neighborhood.”

This precision of wording can be somewhat controlled in translation, but less so in interpretation. Consecutive interpreting, where the interpreter first listens to pieces of the source text, then relays it in the target text rather than speaking in the target text simultaneously, is most widely used in telephonic interpreting and survey research. This type of interpreting requires the interpreter to listen to source language speech first, with a time lag of up to several minutes, before producing utterances in the target language. Psycholinguistics research widely suggests that memory for verbatim linguistic forms lasts only a few seconds (Dam, 1998).

Interpreting implies a linguistic restructuring and retooling, (largely uncontrolled and undocumented), and may create a semantically dissimilar product. In a study evaluating how interpreters interpret based on lexical structure (think questionnaire wording), versus meaning, Dam (1998) showed that 79% of interpreted texts were either partially or completely lexically dissimilar to the source text, even though the source was worded in such a way that it would be clear and comprehensible to speakers of that language, but also could be interpreted in a lexically similar way to the target language. That is, even though they could have been interpreted in a way that would render the result structurally similar to the original, in 79% of cases, they were not. Dam suggests the interpreters used meaning, rather than lexical structure, or a combination of the two, the majority of the time to interpret. What does this mean to the survey researcher?

Given that a questionnaire interpretation is by definition a wording change from the original, one in which not only the words are altered, but in which the social and cultural references are also shifted, what impact could this have on respondent answers, and thus of conclusions drawn about linguistic and cultural subgroups? What impact does the lexical restructuring and reorganizing of questions that interpretation requires have on how the respondents hear questions? How might the variations in interpretation affect respondent data? What steps can the survey researcher take to benefit from the ease, flexibility and low initial cost of interpreters, while still ensuring high data quality?

Towards an interpreting methodology: Recommendations from NORC's Culture Methodology Program

There are many steps that can be implemented to help ensure higher data quality when using interpreters. Specific interpreter methodology will of course depend on the details of the project. For example, if interpreters are needed to conduct brief household screenings for a RDD study, the role of the interpreter may be more similar to a medical interpreter (that is, concerned with the transmission of meaning rather than the precision of administration). For a survey requiring longer interviews with complicated subject matter and skip logic, more precise procedures should be applied.

A well-designed training for both interpreters and interviewers that will work with them is important to maximize consistency of administration. Design a brief interpreter training that covers critical elements of survey research, as well as the specific demands of the project. Interpreters should also be given the full English questionnaire in advance so they can familiarize themselves with the wording of both questions and answer choices, and any skip patterns, as well as any technical language.

NORC's Culture Methodology Program recommends distributing translated "critical terms" to maximize consistency on key points if giving the interpreter a fully translated instrument is not possible. This standardizes the way respondents will hear critical elements and phrases from the questionnaire, and reduces the possibility that one respondent will hear the equivalent of "what work do you do", while another hears "what job do you have?" for important concepts (example from Harkness, 2003). Also standardize whether the name of your organization, and the project, will be translated or

left in English, and if translated, what it is. This helps to keep a coherent image of the project in the field.

Certain terms may best be left untranslated, even with a multilingual sample, in a US context (Sha, 2004). For example, terms such as “condo” may be concepts that even a non-English speaking US inhabitant might know in English, or do not exist clearly in their language.

Quality control procedures for interpreters will vary based on the confidentiality needs of each project. Because of the transient nature of interpreted speech, quality control necessitates a bilingual individual to either listen to the actual interview or to a recording. If the project, and IRB, decide that this is acceptable, the individual can monitor that the interpreter is using the translated terms correctly, is interpreting as precisely as possible without changing the meaning of the questions or answer choices, is interpreting each question and answer choice in its entirety without eliminating anything, and is serving as a conduit for communication between the interviewer and respondent rather than another party in the conversation. Good telephonic interpreting companies do regular quality monitoring of their interpreters, and may be willing to integrate your quality measures with theirs.

However, even the most well-trained and effective interpreter is not working alone. It is essential to train interviewers to work with interpreters. NORC designs and administers special training for the selected interviewers who work with interpreters. Interviewers are trained on the administrative procedures of working with the selected company, but also on how to work with interpreters as survey professionals.

These interviewers are trained to speak at a moderate pace, and pause at appropriate points. Literature on interpretation suggests that as the speed of source speech increases, the accuracy of interpretation declines (Gerver, 1969). This is a seemingly simple training point that can greatly enhance the quality of the questionnaire that the respondent hears. Take for example the following publicly available question in the General Social Survey:

“Now I would like you to think about the people who live in this household. Please include any persons who usually live here but are away temporarily – on business, on vacation, or in a general hospital – and include all babies and small children. Do not include college students who are living away at college, persons stationed away from here in the armed forces, or persons away in institutions. Is everyone in this household related to you in some way?”

If this question is read rapidly and without pause, the likelihood of the interpreter remembering the entire thing precisely and converting it accurately into the target language is low. Depending on the interpreter, the respondent may hear the equivalent of: “who lives with you? Is it only your family?” But, if the interviewer speaks slowly and pauses at appropriate times, it enhances the ability for the interpreter to reformulate

the questionnaire as precisely as possible into the target language, rather than just relaying the perceived meaning.

We also institute an interpreter re-briefing at the start of each call before the respondent is on the line, to refresh the interpreter's memory or in case it has been a while since the interpreter training. Recall that most telephone interpreters are accustomed to calls of 5-10 minutes in duration, where the goal of the call is the transmission of meaning. Therefore, before the respondent is on the line, our interviewers briefly reiterate the unique nature of interpreting for survey research: to be as precise as possible with the interpretation, asking questions that may seem repetitive but are asked for a reason, that sometimes we may ask the respondent to choose only a yes or no answer, or an answer according to a set of response categories.

The interviewer also advises the interpreter of the anticipated length of the call. If the call is long, interpreters are permitted to take breaks, as this type of interpreting in particular is a cognitively demanding task. Since the process of interpretation will likely double the length of the interview, everybody may need bathroom and rest breaks during longer instruments.

Interpreter services can be of use to the survey professional. For certain studies where a full instrument translation is not feasible, or it is not known in advance what languages will be encountered in the field, commercial telephone interpreters in particular can offer a survey greater outreach to limited English speaking respondents, with immediate access, flexibility and low initial cost outlay. However, many of the potentially major data quality pitfalls discussed in the written translation methodology literature apply to using interpreters, and even more that the existing literature cannot address.

In light of increasing interpreter use in multilingual surveys, future research is needed to address the effects of interpretation. What is the impact of variability in interpretation on data quality? Is there an "interpreter effect"? If so, how does it interact with any interviewer effect? Does the increased length of the survey in interpretation raise issues of increased respondent burden, fatigue, or greater non-response? In the meantime, a new interpreter methodology can be implemented which can help projects take advantage of the benefits of interpreters, while avoiding the hazards along the way.

References:

- Bullinger, M., 1998, Translating Health Status Questionnaires and Evaluating Their Quality: The IQOLA Project Approach, *Journal of Clinical Epidemiology* 51(11):913-923.
- Dam, H., 1998, Lexical Similarity vs Lexical Dissimilarity in Consecutive Interpreting: A Product-Oriented Study of Form-Based vs Meaning Based Interpreting, in F. Pöchhacker and M. Shlesinger (eds.), *The Interpreting Studies Reader*. New York: Routledge.
- Gerver, D., 1969, The Effects of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters, in in F. Pöchhacker and M. Shlesinger (eds.), *The Interpreting Studies Reader*. New York: Routledge.
- Harkness, J., 2003, Questionnaire Translation, in J. Harkness, F. van de Vijver, and P. Mohler (eds.), *Cross-Cultural Survey Methods*. New York: Wiley.
- Harkness, J., Pennell, B., and Schoua-Glusberg, A., 2004, Survey Questionnaire Translation and Assessment, in *Methods for Testing and Evaluating Survey Questionnaires*, S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin and E. Singer (eds.), New York: Wiley
- Harkness, J., van de Vijver, F., and Johnson, T., 2003, Questionnaire Design in Comparative Research, in J. Harkness, F. van de Vijver, and P. Mohler (eds.), *Cross-Cultural Survey Methods*. New York: Wiley.
- Rasinski, K., 1989, The Effect of Question Wording on Public Support for Government Spending, *Public Opinion Quarterly*, 53(3): 388-394.
- Raisinski, K., Ernst, M, and Haggerty, C., 2003, Do Small Wording Changes Affect Comparability?, presented at the American Association for Public Opinion Research annual conference, Nashville, TN.
- Shoua-Glusberg, A., 2004, Translation of Survey Instruments, presented at Washington D.C. American Association for Public Opinion Research meeting.
- Sha, M. 2004, Managing Survey Translation: Methodology for Researchers who are not Multi-lingual, presented at the American Association for Public Opinion Research annual conference, Phoenix, AZ.
- Smith, T, 2003, Developing Comparable Questions in Cross-National Surveys, in J. Harkness, F. van de Vijver, and P. Mohler (eds.), *Cross-Cultural Survey Methods*. New York: Wiley.
- Smith, T., 2004, Developing and Evaluating Cross-National Survey Instruments, in *Methods for Testing and Evaluating Survey Questionnaires*, S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin and E. Singer (eds.), New York: Wiley
- Smith, T., 1987, That Which We Call Welfare by Any Other Name Would Smell Sweeter: An Analysis of the Impact of Question Wording on Response Patterns, *Public Opinion Quarterly*, 51(1):75-83.