STUDY SERIES
(Survey Methodology #2012-04)

Evaluating Multilingual Questionnaires:
A Sociolinguistic Perspective

Yuling Pan
Marissa Fond

Center for Survey Measurement
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: December 26, 2012

Evaluating Multilingual Questionnaires: A Sociolinguistic Perspective

Abstract: This study aims to develop an assessment tool to evaluate multilingual questionnaires by categorizing the types of translation issues that can lead to measurement errors in cross-cultural surveys. Based on the results of two multilingual projects that cognitively pretested the 2010 U.S. Census questionnaire in five languages and the American Community Survey questionnaire in two languages, we developed a coding scheme guided by sociolinguistic approaches to language and culture to evaluate translated questionnaires by classifying translation issues in terms of Linguistic Rules, Cultural Norms, and Social Practices. In this paper we discuss how the coding scheme is useful in the evaluation of multilingual questionnaires and how it could be integrated productively into the development of such questionnaires and the early rounds of translation. We also suggest feasible solutions to translation issues, to ensure translation quality and achieve not only semantic but functional equivalence across translations.

Running head: EVALUATING MULTILINGUAL QUESTIONNAIRES

Evaluating Multilingual Questionnaires: A Sociolinguistic Perspective

Yuling Pan and Marissa Fond

Author Note

Yuling Pan, Center for Survey Measurement, U.S. Census Bureau; Marissa Fond, Center for Survey Measurement, U.S. Census Bureau.

Correspondence regarding this article should be addressed to Yuling Pan, U.S. Census Bureau, 4600 Silver Hill Road, Washington, District of Columbia 20233. Email: yuling.pan@census.gov

Evaluating Multilingual Questionnaires: A Sociolinguistic Perspective

Introduction

With globalization facilitated by technology and influenced by the surge of immigration that blurs traditional linguistic and cultural boundaries, government statistical agencies and survey research organizations realize the increasing need to develop multilingual questionnaires in order to conduct survey research. A common practice of developing multilingual questionnaires is to translate the source language questionnaire into target languages. However, reliable translations of source language questionnaires cannot be developed by simply producing translations that are technically accurate (Pan & de la Puente, 2005). Rather, a quality translation incorporates the social, cultural, and linguistic elements of each target language to better match respondents' experience and ensure data quality.

Prior research on survey translation (e.g., Harkness et al., 2003; Pan & de la Puente, 2005; Forsyth et al., 2007) has a tendency to focus on the procedural aspect of the translation-review process, rather than the effectiveness of the result. While the method for developing an accurate translation has been refined extensively through this line of research, little attention is paid to evaluating the final product of this process. This paper demonstrates the challenges in evaluating translated survey questions and argues for the need to develop a translation-review framework guided by sociolinguistic theories. Based on findings from 112 cognitive interviews conducted by the Census Bureau to evaluate the 2010 Census questionnaire in five languages, we developed a coding scheme to classify translation issues[1] that are caused by different linguistic

---

[1] We use the term "issue" throughout this paper to describe aspects of translations that were not successful in conveying the intended meaning from the survey designers. This term encompasses typographical errors, linguistic messages and cultural differences, and respondent errors that relate to the usability of the survey instrument and/or

conventions (usage of words and grammar), cross-cultural communication norms (appropriate expressions of a concept), and social practices (knowledge needed to process a concept or to answer a question). We then applied the coding scheme in a second study to evaluate the Chinese and Korean translations of the American Community Survey. Thus, in this study we aim to determine the types of translation issues that can lead to measurement errors in cross-cultural studies, and we discuss how the coding scheme can be useful in developing multilingual questionnaires. We also explore a number of feasible solutions to address the types of translation issues encountered, so as to ensure translation quality.

Background

Before we present our study, some discussion of various approaches to survey translation is necessary to highlight the need for systematic evaluation of multilingual questionnaires. Typically, the first step in developing multilingual questionnaires is the translation of a questionnaire from a source language into target languages. Because questionnaires are rarely developed with multiple languages in mind, the need for direction on survey translation has been acknowledged by key statistical agencies throughout the world, locally-based organizations (many in the area of health care delivery), and professional researchers in the fields of cross-cultural studies and survey methodology. Over the past few decades, the scholarly community conducting cross-cultural studies has provided useful insights on different approaches to the translation of data collection instruments in multiple languages.[2] The cross-cultural survey

---

format. We believe that "issue" is a more accurate descriptor than "problem" or "mistake" because there is no ascription of blame, to the survey designers or the respondents, in this paper.

[2] Harkness et al. (2003) provide an extensive review of cross-cultural survey methodology. Also, the work of Behling and Law (2000) serves as an example of how the translation of data collection instruments is explicitly

literature describes a number of approaches used to develop questionnaires in multiple

languages, and we will briefly describe two of the most influential here. The first approach is to

adopt the questionnaire to the target language and the second is to *adapt* it.

Adoption calls for the most direct translation of the questionnaire from the source

language to the target language without regard to the linguistic and cultural subtleties that may

impact the intended meaning of the question. This approach is based on what is frequently

referred to as the "Ask-the-Same-Question" model (see Harkness et al., 2003). This model is

based on the often erroneous assumption that a question that seems to be understood by

respondents in the source language will also be equally comprehensible in the target language,

and it ignores semantic and cultural differences that exist across languages.

The second approach, *adaptation*, also uses the source questionnaire as the base, but

allows for components of the survey questions to be modified, in ways independent of the

inevitable changes that result from the translation from the source to the target language, in order

to make the questionnaire interpretable in the target language. This approach to survey

translation attempts to account for semantic, conceptual, and other differences that exist across

languages, and the modifications can take several forms (Harkness et al., 2003), including

terminology explanations, adjustments to language-specific rules, convention-driven differences,

(e.g., writing conventions), and cultural practices and sensitivities associated with the target

language, such as indicators of politeness. The adaptation approach can help to ensure that

survey questions (in both the source and target languages) measure the same or similar construct

and avoid concept biases, and therefore achieve *functional equivalence*; this means that the

instrument conveys the meanings of questions, instructions, and response options in ways that

---

addressed, and other guidance can be found in specific case studies of survey translation, for example, McKay et al. (1996), Potaka and Cochrane (2002), and Schoua-Glusberg (1992).

are comparable across versions. Such questionnaires are more likely to provide reliable,

complete, accurate, and culturally-appropriate information than instruments developed using

other techniques (Behling & Law, 2000; McKay et al., 1996).

The debate on the merits of adoption vs. adaptation provides a conceptual framework

within which to consider survey translation, in that it has led survey researchers to focus on the

intended meaning of a survey question instead of, for example, the syntactic structures of a

question in the source and target languages. This attention to the functional equivalence of

survey questions and the growing popularity of the adaptation approach incited changes to the

actual task of translation, which can be performed using a variety of methods or techniques[3].

Recent literature in cross-cultural studies advocates the use of the "committee approach"

(Schoua-Glusberg, 1992; McKay et al., 1996; Harkness, van de Vijver, & Mohler, 2003; Forsyth

et al., 2007), which is more comprehensive and collaborative because it relies on input from a

team whose members (subject-matter experts, survey designers, and researchers) have skills that

augment those of a translator. An additional step is called for by this approach: the pretesting of

the translated instrument. In this approach, several translators independently translate the

instrument from the source language to the target language. Then the translators, the translation

reviewers, and other members of the team discuss the translated versions of the instrument. A

reconciled version of the translated instrument is produced, and this data collection instrument is

then pretested. After the pretesting is complete, the adjudicator and other committee members

convene again to finalize the instrument.

---

[3] Some of the most commonly used techniques are "simple direct translation," "modified direct translation," "back translation," and "committee approach." See Harkness et al. (2003), Behling and Law (2000), and Pan and de la Puente (2005) for details.

Taking the view that the task of translating a data collection instrument is a process that entails the participation and cooperation of a number of individuals with complementary skill sets and professional experience, the U.S. Census Bureau Translation Guidelines recommend the committee approach (Pan & de la Puente, 2005). They require that every translation team assigned to produce final versions of Census Bureau translated questionnaires and supporting materials should involve five different groups of professionals: translators, reviewers, subject matter experts, survey methodologists, and adjudicators. In addition, all translated instruments and materials should be pretested with speakers of the target languages.

To summarize, prior research on survey translation has focused primarily on the translation process and techniques for the performance of translation, rather than the evaluation of the result. While most agree that achieving functional equivalence is the goal of survey translation, achieving this goal can be deceptively complicated, particularly without guidelines to measure it. Many questions remain regarding functional equivalence and the success of the translation process, such as how functional equivalence can be measured and how the quality of translated material can be consistently and reliably evaluated. Levin et al. (2009), in a thorough review of work on cross-cultural cognitive interviewing methods for pretesting surveys, note that the range of issues identified in multilingual survey pretesting is extensive (p. 14). They summarize six classification systems that researchers have used to explain the problems that were uncovered in translation pretesting, and while the systems differ in the descriptions or operationalizations of each category, they all include references to linguistic or "translation" issues, issues that are "culturally" based, and issues with survey navigation generally (see Carrasco, 2003; Schoua-Glusberg, 2006; Goerman & Caspar, 2007; Harkness, Mohler, & van de Vijver, 2003; Willis et al., 2008; Willis & Zahnd, 2007).

While these research efforts began to foment a more systematic examination of translation issues, there is still a need for further empirical study of the evaluation of translation quality in multiple languages. Survey project managers and translation review committees still face the challenge of how to systematically assess translation quality, particularly in a language that they do not speak or in which they do not have expertise. They are still faced with four basic questions: 1) How can we determine how successful a translation is? 2) When translation issues are identified in the expert review or cognitive testing process, how can we best articulate and describe the nature of the problems? 3) Can patterns of inadequacies in translated materials be identified so that they can be anticipated and addressed efficiently? 4) What are the feasible and effective solutions to the problems identified?

Project 1: Towards the Identification of Functional Equivalence

The research described in this paper is an attempt to fill in the aforementioned knowledge gap and to further research in survey translation. We aim to develop a method to categorize translation issues and to systematically evaluate translated questionnaires, by addressing the question of functional equivalence from the perspective of the relationship between linguistic code and social/cultural context. Our approach differs from those reviewed above (Levin et al., 2009) in that we ground our analysis and coding of translation issues in the theoretical framework of sociolinguistics. Instead of considering language to be a category separate from cultural and societal influences and concerns, we examine the different facets of language, and language use, that affect the utility of a survey instrument.

A sociolinguistic perspective on translation quality requires that a sound, effective survey translation function at micro- and macrolinguistic levels. The microlinguistic level concerns the

word choice and sentence structure in a translation, meaning that the translation should consist of accurate, appropriate wording and use the correct terminology to convey the meaning of each individual word in the source text. Moreover, the translation should follow the grammatical structure of the target language, and sound natural to a native speaker of that language. The macrolinguistic level is oftentimes referred to as the pragmatic level of translation (Pan et al., 2007), which means that a translation should achieve its intended function (as in "functional equivalence"), in that respondents who read the translated questions or survey materials can understand the intended meaning and take necessary actions as a result to provide the data requested.

When we talk about translation issues at the pragmatic level, we talk about the frames of reference, or schemata, that people rely on for interpreting a translated item. This involves the sociocultural context as well as background knowledge or experience and communication norms common among speakers of the target language. In order to ensure that translated questionnaires are appropriate at the pragmatic or "functional" level as well as the microlinguistic level of lexicon and syntax, we need to consider one basic principle in sociolinguistics: that a language is inseparable from the culture and society in which it is used (Gumperz, 1999; Holmes, 1992; Tannen, 2005). Language use inevitably reflects, and perpetuates, the values and social practices of a given culture. In order to tackle problems in translation, we need to analyze not only linguistic rules that govern the sentence structure or word order of a specific language, but also the cultural norms of expressing certain concepts and the social practices encoded in linguistic expressions. Therefore, there are three components that we need to consider in our analysis: *Linguistic Rules*, *Cultural Norms*, and *Social Practices*. Linguistic Rules refer to language-specific rules, such as the grammar, the word order, or internal sentence structure of a language.

These issues are identified at the word or clause level. Cultural Norms refer to the ways of doing

certain things in a given culture, such as communication style, the discourse sequences for

presenting information, and culture-specific ways of showing politeness. This category allows us

to examine language beyond the boundary of the sentence, and how a communicative event

unfolds (e.g., the question-answer rhythm of a survey). Social Practices refer to daily or

institutionalized practices in a society, including social institutions, educational systems, or

personal experiences as influenced by culture and society. For example, survey interviewing may

be a common practice in American culture, but it could be a foreign concept to people who are

from China or Vietnam (see Pan, 2008; Pan et al., 2009). Because language encodes the cultural

values and salient social practices of a particular cultural group, and language use is always a

reflection of cultural norms and social knowledge (Gumperz, 2001), the three components of

Linguistic Rules, Cultural Norms, and Social Practices can serve as the guiding principles for us

to evaluate the quality of a translation.

Development of the Coding Scheme: Method

The coding scheme was based on findings from two cognitive testing research projects:

1) the 2010 Census questionnaire in five languages (English, Chinese, Korean, Vietnamese, and

Russian), and 2) the American Community Survey questionnaire in two languages (Chinese and

Korean). The first project involved the 2010 Census questionnaire, a self-administered survey

with 10 basic demographic questions. First, the entire questionnaire was translated from English

into the four target languages. Then, a total of 112 cognitive interviews were conducted with

monolingual speakers of the five language groups. Cognitive interviewing, a method referenced

in the pretesting literature above, is a semi-structured interview method used to "study the

manner in which targeted audiences understand, mentally process, and respond to the materials"

provided by survey researchers (Willis, 2005, p. 3). Based on the summary of findings from this

first project, we developed the coding scheme to code the translation problems observed. In the

second cognitive testing project, evaluating the Chinese and Korean translations of the American

Community Survey questionnaire, we applied this coding scheme to evaluate the translations

with the goal of identifying possible solutions to translation issues.

The methodology in the cognitive testing procedures for the two projects involved the

following steps that ensured a systematic evaluation of the translated questionnaires: 1) a team of

three language experts (native speakers with experience in cross-cultural methodology and

cognitive testing) was assembled for each language, and these team members went through a

two-day training program in cognitive interviewing and project-specific requirements; 2) using a

committee approach, the language teams translated the cognitive interview protocol, which was

developed by Census Bureau survey methodologists, from English into the target languages; 3)

the language teams conducted cognitive interviews with respondents with a range of

demographic characteristics who spoke the target language; 4) after the first round of interviews,

the language teams met to identify problematic translations, suggest alternative translations, and

articulate their justifications for their suggestions, based on the findings of the interviews. The

suggested changes were tested in the second round of interviews; 5) finally, after all of the

cognitive interviews across the languages were completed and summarized, the language teams

met once more to reassess the translation of the census questionnaire, and suggested final

recommendations for alternative translations, based on the outcomes of the interviews (Pan et al.,

2007).

After the cognitive testing was completed, we analyzed the results documented by the language experts. In the cognitive interview summaries produced for each interview, each issue uncovered was noted and explained by the language experts, who then offered recommendations for possible solutions based on their linguistic and cultural knowledge. Working with the explanations of the issues and the suggested solutions, it became apparent that while most translated census questions and instructions could be considered correct and accurate in terms of grammar and word usage, respondents demonstrated difficulties understanding the intended meaning of questions and providing answers that were satisfactory to them. We then classified each difficulty according to whether the issue was due to Linguistic Rules, Cultural Norms, or Social Practices, noting the characteristics that each issue had in common and the rationale for grouping it with similar issues. In addition, we noticed that there were some issues that arose from production errors like typographical errors or accidentally-omitted words. We also noticed that some issues were caused by respondents not paying attention and making mistakes in answering the questions. To address these issues, we added the following two categories: Production Errors and Respondent Errors. We introduce these as separate from the categories of Linguistic Rules, Cultural Norms, and Social Practices because they are qualitatively different; they represent two types of minor errors or "mistakes" that are easily corrected with a second pair of eyes, or a second review. For example, a misspelled word that is considered by language experts to be a clear typographical error, and not a possible alternative spelling, would be grouped under Production Errors. A respondent who answered a question incorrectly because they omitted a word while reading the question would have committed a Respondent Error. In this situation, if the respondent is asked to read the question again, they do so without any errors. We found it necessary to record these categories of errors, rather than dismiss them outright as

easily remedied, because in the event that we found more Production or Respondent Errors in a particular section of the survey, within a single translation or across multiple languages, these issues would be worthwhile to address in usability testing.

The Coding Scheme

The completed coding scheme includes the three main categories listed above that reflect the components of Linguistic Rules, Cultural Norms, and Social Practices, as well as Production Errors and Respondent Errors. We briefly summarize the categories here.

Table 1. *The Coding Scheme*

Codes	Explanations
Linguistic Rules	This category classifies issues in a translation, at or below the level of a sentence, that are due to lexical items, morphology, syntax, usage conventions, etc.
Cultural Norms	This category refers to issues that arise when concepts that are expressed one way in English (the source language of the survey) are expressed in a different way in the target language (e.g., personal address conventions, numbering/counting, kinship terms, time references, conversational norms, etc.).
Social Practices	This category classifies issues with concepts that can be described in English but cannot be translated into the target language because either the concept does not exist in that culture, or respondents have no experience with the concept.
Production Errors	This category refers to survey production problems that are simple mistakes (e.g., word omissions, typographical errors) that can be easily corrected.

Respondent Errors	This code refers to actions taken by respondents while reading or answering a questionnaire that they themselves identify to be easily-corrected mistakes (e.g., those caused by inattention).

How the coding scheme works

Through the examples of issues uncovered in the review of the pretesting of the 2010 Census questionnaire that follow in this section, we demonstrate how the coding scheme works. In order to reach the goal of producing appropriate and accurate translations, it is necessary to understand the unique properties of each individual issue that arises in a given language, in a given questionnaire. These examples are presented for the purpose of demonstrating how we coded translation issues, based on cognitive interview summaries from the language teams; they are intended to be illustrative, not exhaustive.

Coding of Linguistic Rules issues

Based on cognitive interview summaries, we identified evidence that indicated problems with respondents' understanding of the translated questions due to the subtle differences in word usage or sentence structure between the target language and the English language rules. An example to illustrate this follows.

Figure 1: Question 2 from the 2010 Census.

Regarding this question, there were translation issues in three of the four languages that were classified as related to Linguistic Rules. In Korean, "live-in baby sitter," seen here as part of the third response option, was translated as "always-staying baby sitter," which sounded as awkward, and confusing, in Korean as it does in English, indicating that the lexical items chosen to translate the term were not optimal; while a respondent could perhaps guess what "always-staying baby sitter" might mean, the phrase does not convey its intended meaning as clearly as possible. In Vietnamese, the translation of the question included repetitive auxiliary verbs which obscured the meaning of the question, in particular the reference time; the verbs needed to be deleted to make the sentence sound natural and the meaning clearer in Vietnamese. These are all issues classified as Linguistic Rules issues because they implicate lexical items and syntax; the resulting awkwardness of the translations was due to the failure to adapt the English original to the linguistic features of the target language.

It is important to bear in mind that Linguistic Rules issues are not necessarily errors. The translations may be grammatically correct, but they are not natural in the target language, in that the translations sound to respondents like translations, rather than materials written in their language. Put another way, these issues are examples of words or phrases that respondents would not be likely to hear or read in their language. These issues can hinder respondents' comprehension of the translated questions and answer categories.

Coding of Cultural Norms issues

The problems in this category usually resulted from the different ways of expressing a similar concept in different cultures. This means that a concept included in the English questionnaire exists in the target culture, but due to different conventions of expression or a different focus on certain values in American culture and the target culture, the concept is conveyed differently through language. For another example, see Question 7 in Figure 2:

Figure 2: Question 7 from the 2010 Census.

7. What is Person 1's age and what is Person 1's date of birth?
Please report babies as age 0 when the child is less than 1 year old.
Print numbers in boxes.
Age on April 1, 2010   Month   Day   Year of birth

An example of a Cultural Norms issue was present in the translation of this question into Korean. More than two thirds of Korean speakers experienced confusion writing their age because the Korean convention of counting age is different from the American one. In Korean culture, newborn babies are considered one year old, so someone who would be 50 years old as Americans count age would be considered 51 by Korean speakers. In order to address this issue, the translation must specify that the form is asking about the American way of indicating age. This is a Cultural Norms issue because the Korean speakers' interpretation arises not from the lexical items used, or the syntax of the translation, but from the cultural background of the speakers; this means that when they respond to the question, they are answering based on an interpretation of the question that was not intended by the American English-speaking survey designers.

Coding of Social Practices issues

There were many examples of questions on the 2010 Census questionnaire that gave rise to Social Practices issues. In Question 2 cited above in Figure 1, the term "foster children" proved problematic for Chinese, Korean, and Vietnamese speakers. There are existing terms for "foster children" in these languages, and so these terms were used in the translations, but the terms refer to concepts that are quite different from the concept of "foster children" in the United States. In these languages, the terms for "foster children" were understood as meaning "children under the temporary care of relatives or friends." The concept of a foster program administered and supported by the government was not retained in the translation. This type of program does not exist in China, Korea, or Vietnam, so there is no lexical item in these languages that can be used to describe it; therefore, a more descriptive phrase that provides additional information is necessary.

Another example of a concept that appeared to be uniquely American is found in Question 3, seen here in Figure 3.

Figure 3: Question 3 from the 2010 Census.

Question 3 provides response options relating to the ownership of the residence, but the question itself proved to be difficult. Speakers of all four tested languages were unsure about what a "mobile home" was, as mobile homes are not common in other countries (and in fact, speakers of other languages such as Spanish will often adopt the American word "trailer" rather than describe the concept in their native language). A final example of a concept that could not be

readily translated into another language clearly and unambiguously was "nursing home," found

in the response options for Question 10 as seen in Figure 4:

Figure 4: Question 10 from the 2010 Census.

For the term "nursing home," the approximate translations were found to be inaccurate (or

inadequate). The target languages have equivalent terms for "nursing home," but the meanings

are different from the American concept. In Chinese, the translation meant, to various speakers,

either a mental hospital (to Hong Kong Chinese) or a recreational resort (to Mainland Chinese);

in Korean the translation meant a resting place; in Russian it meant a medical establishment; and

in Vietnamese it meant a luxurious resort. Given that these cultures do not have experience with

American-style nursing homes, the translations were unable to capture that meaning. A longer

and more descriptive phrase is necessary in order to make sure that the individual filling out the

questionnaire understands the question as it is intended.

Identifying overall patterns of problems

Once all of these issues were identified and categorized according to the coding scheme,

it was possible to quantify them in order to determine which types of issues were most

problematic in the 2010 Census translations. The chart below (Figure 5) summarizes the findings

of types of issues identified in the four target languages (excluding the English source material)

during the cognitive testing process. There are four main types of issues coded: Linguistic Rules,

Cultural Norms, Social Practices, and Production Errors. We did not find any Respondent Errors

in this phase of testing, because the analysis was conducted using a composite summary of the

issues in each language, rather than summaries of individual interviews with respondents. Any

respondent difficulties were likely not significant enough to be included in the summary report;

this may be explained by the fact that the 2010 Census questionnaire contains only ten

demographic questions, and it is relatively simple in terms of form navigation.

Regarding these four types of issues, it quickly became apparent that Linguistic Rules

and Cultural Norms issues were the most common, as they comprised 33% and 39% of the total

163 issues discovered, respectively. The remaining issues were divided between Social Practices

and Production Errors, with 18% and 10% respectively.

Figure 5: Types of issues in all four languages for the 2010 Census questionnaire.
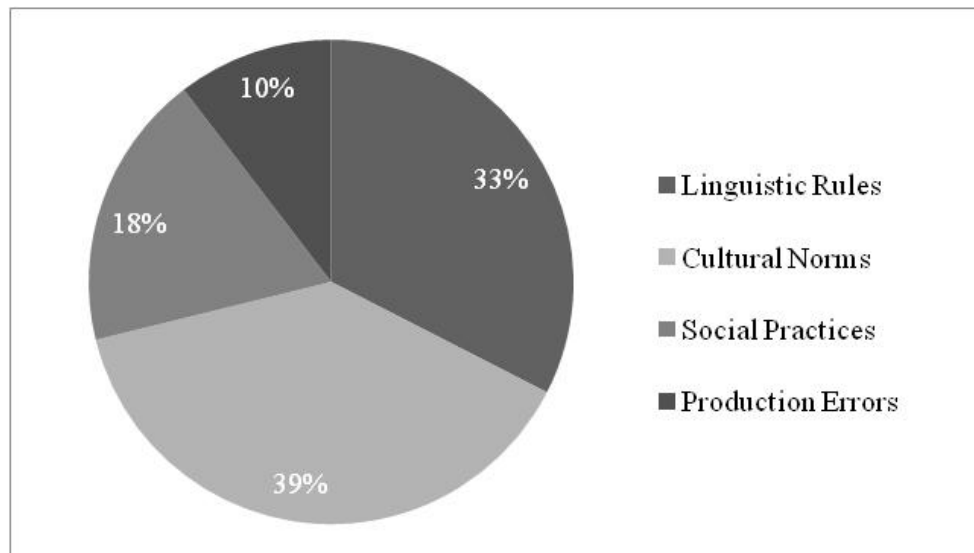
Next, in order to deepen our understanding of the translation issues and how the coding

scheme works, we wanted to determine whether each of the four target languages seems to have

the same pattern of issues identified, or not. The chart below (Figure 6) shows that Linguistic

Rules and Cultural Norms issues constitute the majority of issues in each language, as we already

discussed. Korean and Vietnamese show similar proportions, but in Chinese, the Linguistic

Rules category comprises a much larger percentage of total errors.

Figure 6: Issues as a % of total (per language) in the 2010 Census questionnaire.
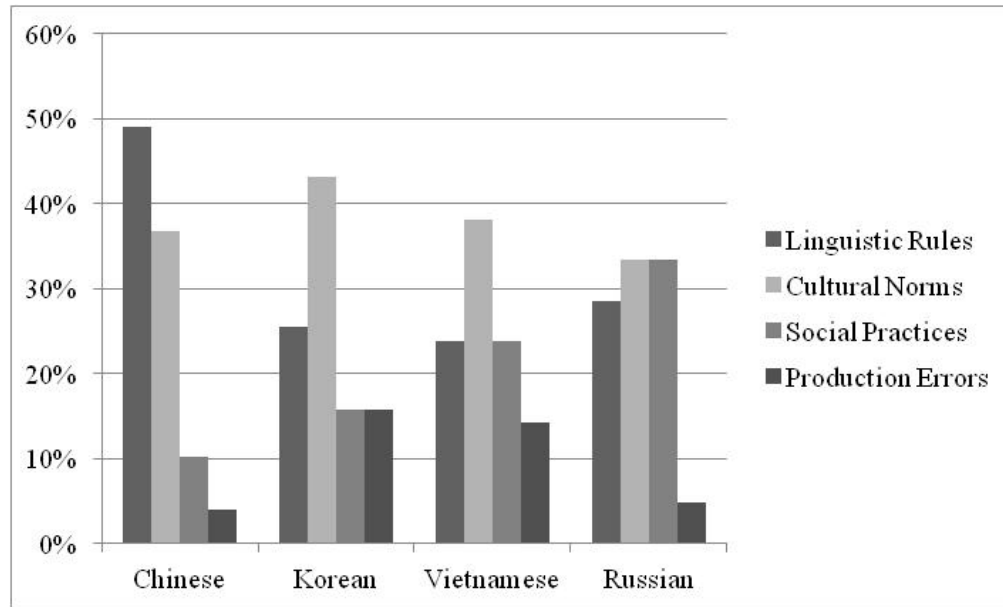
We can use this information to focus our attention on explaining why this might be; the

proportion of Linguistic Rules issues might be much higher in Chinese than it is in the other

languages because the Chinese translation had an unusually high number of issues with syntax;

there were many examples of overly complex sentence structure that were not present in the

translations in the other languages (in particular, structures that were patterned closely after the

English original rather than adjusted to typical Chinese structures). Also, while it may appear

that the Russian translation generated more Social Practices issues than the other translations, in

fact the Russian translation had most of the same Social Practices issues as the other three

translations; these issues merely represented a larger proportion of the total number of issues in

the translation as a whole.

This project showed that the coding scheme worked well not only to identify what types of causes were at the root of translation issues, but to show how the translations into different languages contained different proportions of problems (indicating that the translation teams had different skill levels) as well as some similarities (indicating that there might be elements of the survey, designed with English speakers in mind, that are difficult to translate effectively).

Project 2: Towards Possible Solutions for Translation Problems

After the coding scheme was developed using data from the pretesting of the 2010 Census questionnaire, the coding scheme was applied to a different survey pretesting endeavour (a study of the American Community Survey) in order to examine how the scheme worked in a more detailed analysis of cognitive interview data, and to explore possible solutions for the translation issues uncovered.

Using the Coding Scheme with the American Community Survey

In this phase of the project, we applied the coding scheme to the cognitive testing of the Chinese and Korean translations of the American Community Survey (ACS) questionnaire. The ACS is the largest general survey conducted by the Census Bureau, and it covers many topics, such as demographic characteristics, housing, health insurance, education, income, and transportation. The cognitive testing project reported here was conducted as part of a larger project undertaken by the Census Bureau to evaluate the many subsections of the translated ACS questionnaire (which is too extensive to be thoroughly reviewed in one round of cognitive testing and analysis of results). The portion tested for this study included all of the instructions on how

to complete the questionnaire, basic demographic questions, and the section of questions on housing characteristics.

Similar to the pretesting of the 2010 Census questionnaire, the ACS Chinese and Korean pretesting project followed the same methodology for carrying out the cognitive interviews and analysis. A total of 41 interviews were conducted (19 in Chinese, 22 in Korean). The analysis procedure for both multilingual projects was similar, but there were a few differences. For the 2010 Census questionnaire project, the results from the cognitive interviews, which were written up in a summary report by the language teams, were analyzed by Census Bureau researchers. Each issue uncovered in the interviews was noted and explained by the language experts, who then offered recommendations for possible solutions, based on their linguistic and cultural knowledge. Working with the explanations of the issues and the suggested solutions, Census Bureau researchers coded each issue. This review process was important to developing descriptions of each of these categories, and determining how clearly the issues fit into one or more of them; the coding scheme was operationalized, with sharper divisions between categories, as a result of this process.

The analysis of the ACS interviews proceeded slightly differently. Because the coding scheme was ready to be used after the completion of the 2010 Census project, language experts were able to use it while they were administering the protocol to each respondent and while writing up their summaries of the interviews. Therefore, the respondents' answers to each cognitive probe within the protocol were coded by the language expert, who also provided an explanation of the code chosen. Then, a Census Bureau researcher reviewed each cognitive interview summary, as well as the codes and justifications for each issue, and confirmed the assessment (or in some cases, queried the code if the explanation did not seem to match the code

assigned). In this way, each issue was reviewed by two coders: one language expert and one

research analyst.

Uncovering Overall Patterns of Translation Issues

In applying the coding scheme to evaluate the Chinese and Korean translations of the

ACS questionnaire, we see the distribution of the types of issues that were uncovered in the 41

cognitive interviews in Figure 7.

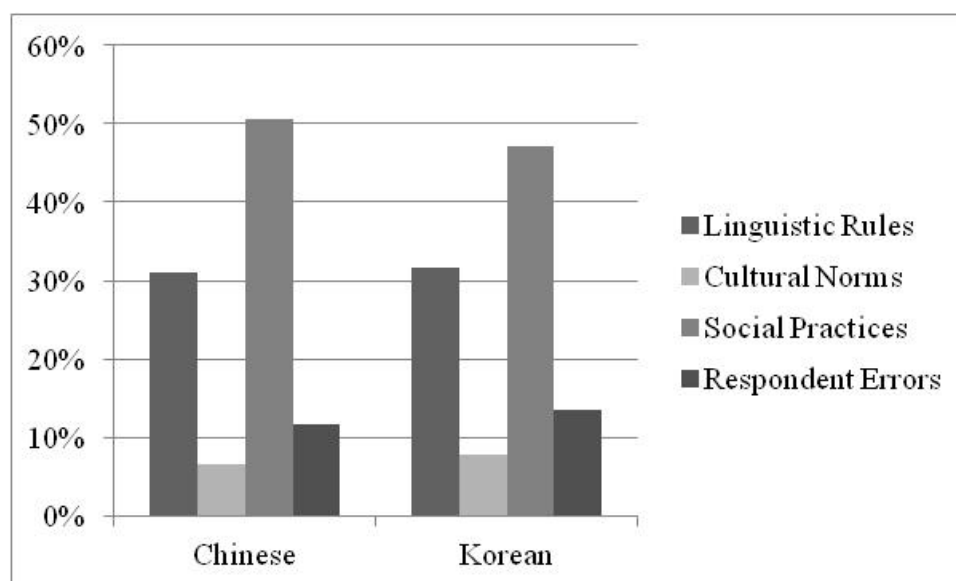*Figure 7:* Types of issues in Chinese and Korean in the ACS questionnaire.
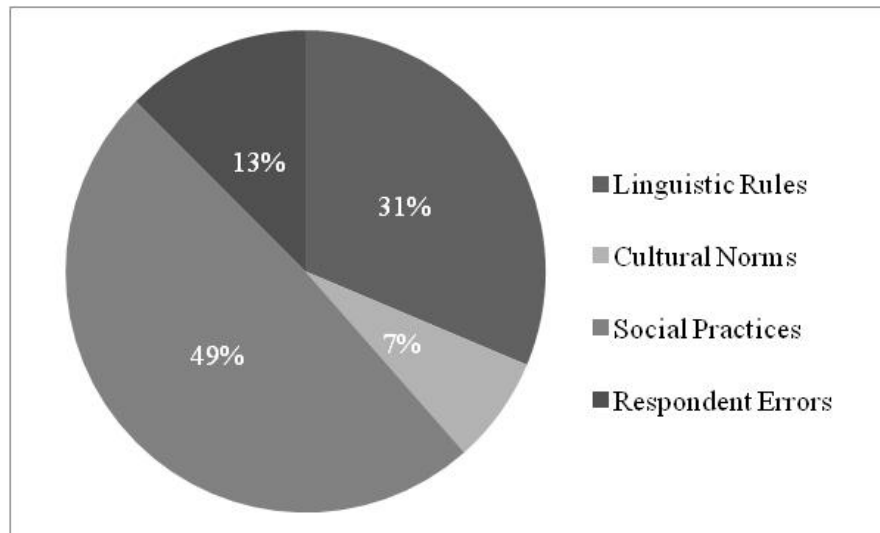
From this chart, we can immediately notice that in this ACS study, Social Practices issues

constitute the majority of the issues (49%) followed by the Linguistic Rules issues (31%). The

predominance of Social Practices issues is not unexpected, because the translations used in this

study had undergone numerous rounds of translation review before the cognitive interviews were

conducted. Even when translation protocols are followed closely, Social Practices issues are the

most difficult to resolve through translation reviews, so it is not surprising that most of the

problems found involved these issues. As for the Linguistic Rules issues, respondent comments

indicated that these translations tended to use long and complex sentences that mirror the original

English sentences without taking into consideration Chinese or Korean language-specific

structures, and employed some outdated as well as high-register terms (e.g., terms familiar to

highly-educated respondents). No Production Errors were identified in the cognitive testing

process, due to the fact that these translations were very carefully reviewed numerous times prior

to the cognitive testing, in contrast to the 2010 Census questionnaire translations. Small

proportions of Respondent Errors (13%) and Cultural Norms issues (7%) were identified.

   In the following chart (Figure 8), we can see the issues as a percentage of the total

number of issues uncovered in each language for the ACS questionnaire translation.

Figure 8: Issues as a % of total (per language) in the ACS questionnaire.

In this chart, it is clear that the proportions of issues in Chinese and Korean are similar. This was

expected because the translations had gone through expert review many times and the

differences in quality among translation teams (which can be significant) were tempered; also,

most of the issues that remain are Social Practices issues common to both Chinese and Korean

societies (which do not have simple translation fixes) and residual Linguistic Rules issues (e.g.,

complex sentences or vocabulary that respondents interpreted differently from the translator's intention).

Trying to uncover patterns in respondents' comments during cognitive interviews is important because if we see that multiple respondents interpreted a question in a way that was not intended by the survey designers, then we know that there is a serious deficiency in the translation. Also, if we see that the respondents interpreted the question in a way that was not intended because of the same general reason, then this not only makes solving the problem easier, but it adds to the bank of information that a survey research organization has about what works in survey questions.
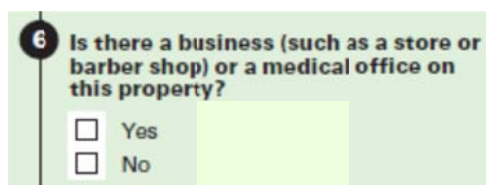
This is why the ability to quantify our results is so important. We are able to see, in one glance, where the most severe problems lie. We can identify the scope of the problem in a given language, or in a given question across languages, allowing us to flag problematic questions for follow-up, as well as to determine if the translation issue was unique to a single language or culture or if it might even be related to the English original (because the English versions are constantly being revised as a result of cognitive interview data). The quantitative results are an important additional tool to aid in translation review; they are not significant in themselves. This is to say that for survey methodologists working on cross-cultural surveys, it is not necessary to know whether, for example, the difference between the number of Linguistic Rules issues and Social Practices issues in the Chinese version of the ACS was statistically significant, but it is interesting to know which questions had the most Social Practices issues so that they can be modified.

Resolving Translation Issues

In addition to identifying issues, the coding scheme can also help to expedite the process of fixing questions that proved to be problematic. For example, Production Errors require careful review and edit, but they do not need additional cognitive pretesting or new translations. A team-based approach to translation review and careful review procedures involving multiple parties are important for catching these problems early in the process.

For Linguistic Rules issues, this is the category that requires a close examination of the subtle meanings associated with a term and attention to whether the translation sounds natural, in the sense that it uses the syntactic structure of the target language instead of that of English. Oftentimes, translations tend to follow the English sentence structure too closely, which can create unnecessary comprehension difficulties for non-English speakers. Following is an example of the Chinese translation of Question 6 in the ACS questionnaire which has two issues in the Linguistic Rules category.

Figure 9: Question 6 from the ACS questionnaire.

In this translation, there are two issues with Linguistic Rules: the translated term for "property" and the term for "business." Both terms are correctly translated in Chinese, in that there are no typographical errors, but the respondents did not get the intended meaning. The translated term "property" was interpreted by Chinese respondents as "real estate property for sale" or "realtor's office." The translated term "business" could imply a large-scale business, like

a Wal-Mart. Therefore, the two terms were coded as Linguistic Rules issues. Based on these findings, the Chinese language team suggested another term meaning "housing unit and yard land" to translate the English term "property" to lessen the confusion. For the term "business," the Chinese language team suggested restructuring the translation to say "*do* business" instead of "*a* business." This change required restructuring the question's syntax and using a verb "do business" instead of a noun phrase "a business." This solution fixed the problem, and the new structured sentence sounded natural in Chinese, based on the language experts' opinion and the respondents' feedback. This example demonstrates that classifying translation issues through the coding scheme gives a clearer direction of how to find solutions, and the solutions become feasible and more systematic.

For Cultural Norms issues, it is necessary to ask whether a certain concept exists in the target culture, and if it does, how it is expressed. These questions will help to identify early on what the cause of the translation difficulty is. The goal is to identify and use culturally-appropriate expressions, employ culture-specific communication styles, translate the discourse structure of English into that of the target language, and incorporate politeness strategies where appropriate. For example, due to the different politeness practices between American and Korean cultures, the Korean translation of the instructions in the ACS questionnaire was found to be too direct, which violated the politeness norms of Korean culture (cf. Pan, 2011; Pan et al., 2010). In one instruction, seen here in Figure 10 incorporated into the "yes" response option, the English wording is: "For renters, answer only if you pay the condominium fee in addition to your rent."

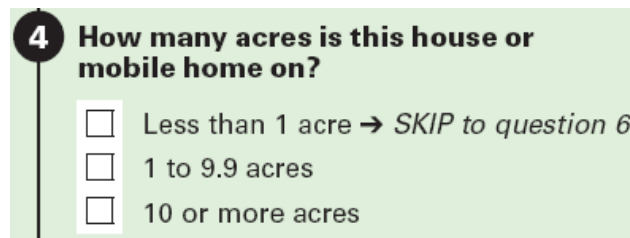*Figure 10:* Question 13 from the ACS housing section.

The original Korean translation of this instruction reads: "Answer only if you rented and pay

money in addition to your rent." This translation was correct and followed the politeness norms

in the English original by using a direct expression of a command. Based on reactions from the

Korean respondents in cognitive interviews, this was coded as an issue in the Cultural Norms

category for a lack of appropriate polite expressions and awkward discourse structure. The

revised Korean translation becomes: "If you are renting a condominium, please write the amount

of condominium fee that you pay in addition to your rent." The revised translation took into

account the necessary expression of politeness ("please") and re-structured the sentence so it

flowed better in the Korean language (by placing the "if" clause at the beginning, before the

command).

While issues in Linguistic Rules and Cultural Norms categories can hinder

comprehension and increase difficulty for respondents, issues in the Social Practices category

indicate that a construct being measured in the questionnaire may be a foreign concept to the

target population. This can very likely lead to measurement errors or item non-response. For

Social Practices issues, it is important to ask if the translated question might measure a concept

or experience that respondents have no knowledge of. If a certain concept or practice does not

exist in the target culture, how can the concept or practice be translated, and how can

respondents come to understand the new concept quickly and clearly? We may need to think of

creative, descriptive ways to translate the concept, or it may be necessary to revisit the source

materials or source questionnaires to collect as much background and contextual information as

possible. From there, translations can include explanations, examples, or notes, plus clear

instructions, which are culturally appropriate and helpful to speakers of target languages. Also,

we recommend flexibility when translating Social Practices issues, and we encourage the use of

descriptive phrases instead of existing terminology.

As an example, see the ACS Question 4 about land measurement (Figure 11):

Figure 11: Question 4 from the ACS housing section.

The land measure "acre" is not used in Chinese- or Korean-speaking countries, so "acre" is not a

salient concept. To overcome this Social Practices issue, the solution was to add supporting

information appropriate for each target language. In Chinese, a note was included that read "One

acre is about 4,000 square meters" (as meters are commonly used to describe land area in

Chinese). In Korean, the supporting note read "One acre is about 1,230 pyeong" (as the

measurement unit "pyeong" is unique to Korean, and commonly used).

Conclusion

The problem faced in the survey translation process is the evaluation of translated surveys in terms of functional equivalence. A translation is "successful" when it is functionally equivalent to the source questionnaire, and this means that not only is the information presented accurately, but that it is understood as intended. The translation process used in survey research so far can address the former, but not always the latter. The latter can be determined through cognitive testing with respondents so that survey methodologists can understand better how their questions are being interpreted; the next step is to aggregate those results so that a translation can be evaluated at a glance, by individuals who may not speak the language(s) into which the survey was translated.

However, we need a way to describe the results of the cognitive interviews that takes all insights from all interviews conducted into account. It is not useful to have dozens of cognitive interviews summarized; they must be interpreted and translated into actionable problems. This is what the coding scheme does. Based on the sociolinguistic approach to language in context, the coding scheme allows us to produce a clear, articulated description of what problems are observed in a translation and what the causes or roots of the problems are so that they can be addressed. Our examples show this; rather than reviewing each one as an isolated event, we can see what questions seem to be inadequate, and see where the issues lie, for each language.

Beyond individual issues or examples, there are also more global solutions to these translation issues. One possibility is to train translators and reviewers to identify different types of issues (Linguistic Rules, Cultural Norms, Social Practices, Production Errors) so that they can be addressed more swiftly. These findings can be shared with survey questionnaire designers and sponsors where appropriate, so that the original surveys can be constructed in a way that makes

them more easily adaptable to other languages. Also, it would be useful to develop a bank of terms and concepts that are commonly used but difficult to translate and to include tested solutions so that future issues can be avoided.

This study builds on an established tradition of coding problems in cross-linguistic surveys (see Levin et al., 2009). A common theme in these coding schemes is the separation of language/translation, culture, and survey navigation issues; the coding scheme outlined in this paper is unique in that it takes a multifaceted approach to the construct of "language," breaking it down into lexical, syntactic, and pragmatic levels and considering the text of the survey in its sociocultural context. By doing so, it offers solutions to these issues based on these principles.

Our next step will be to refine the coding scheme, particularly the way in which we quantify the results. With a larger data set, we can link respondent characteristics (e.g., age, gender, education level) to questions, to determine which segments of the respondent population have particular trouble, and target solutions to their needs. Also, we will refine the coding scheme to make it even simpler to use, to implement a committee approach to coding (to find out where the areas of contention are, if any), to determine inter-rater reliability, and to adjust the coding scheme as necessary based on future cognitive interview data. The anticipated applications of this coding scheme are to better evaluate the quality of translated material, measure how quality improves over time, with training, etc., and identify which questions or instructions are most problematic.

References

Behling, O., & Law, K. S. (2000). *Translating questionnaires and other research instruments: Problems and solutions*. London: Sage.

Carrasco, L. (2003). *The American Community Survey (ACS) en Español: Using cognitive interviews to test the functional equivalency of questionnaire translations* (Statistical Research Division Research Report Series, Survey Methodology #2003-17). Washington, DC: U.S. Census Bureau.

Forsyth, B., Kudela, M. S., Levin, K., Lawrence, D., & Willis, G. (2007). Methods for translating an English-language survey questionnaire on tobacco use into Mandarin, Cantonese, Korean, and Vietnamese. *Field Methods, 19,* 264-283.

Goerman, P., & Caspar, R. (2007). A new methodology for the cognitive testing of translated materials: Testing the source version as a basis for comparison. *Proceedings of the Joint Statistical Meetings (Section on Survey Research Methods),* 3949-3956.

Gumperz, J. J. (1999). On interactional sociolinguistic method. In S. Sarangi, & C. Roberts (Eds.), *Talk, work and institutional order: Discourse in medical, mediation and management settings* (pp. 453-471). Berlin: Mouton de Gruyter.

Gumperz, J. J. (2001). Interactional sociolinguistics: A personal perspective. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 215-249). Malden, MA: Blackwell Publishing.

Harkness, J. A., Mohler, P. P., & van de Vijver, F. J. R. (2003). Comparative research. In J. A. Harkness, F. J. R. van de Vivjer, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 3-16). Hoboken, NJ: John Wiley & Sons.

Harkness, J. A., van de Vivjer, F. J. R., & Mohler, P. P. (Eds.). (2003). *Cross-cultural survey methods*. Hoboken, NJ: John Wiley & Sons.

Holmes, J. (1992). *An Introduction to Sociolinguistics*. London: Longman Group.

Levin, K., Willis, G. B., Forsyth, B. H., Norberg, A., Stapleton Kudela, M., Stark, D., & Thompson, F. E. (2009). Using Cognitive Interviews to Evaluate the Spanish-Language Translation of a Dietary Questionnaire. *Survey Research Methods, 3,* 13-25.

McKay, B., Breslow, M.J., Sangster, R.L., Gabbard, S.M., Reynolds, R.W., Nakamoto, J.M. & Tarnai, J. (1996). Translating Survey Questionnaires: Lessons Learned. *New Directions for Evaluation, 70,* 93-104.

Pan, Y. (2008). "Cross-cultural Communication Norms and Survey Interviews." In Sun, H. and Kádár, D. (eds.), *It's the Dragon's Turn – Chinese Institutional Discourse(s)*, 17-76. Berne: Peter Lang.

Pan, Y. (2011). Methodological Issues in East Asian Politeness Research. In D. Kádár & S. Mills (Eds.), *Politeness in East Asia: Theory and practice* (pp. 71-97). Cambridge: Cambridge University Press.

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* (Statistical Research Division Research Report Series, Survey Methodology #2005-06). Washington, DC: U.S. Census Bureau.

Pan, Y., Landreth, A., Hinsdale, M., Park, H., & Schoua-Glusberg, A. (2007). Methodology for Cognitive Testing of Translations in Multiple Languages. *Proceedings of the Joint Statistical Meetings (Section on Survey Research Methods).*

Pan, Y., Sha, M., Schoua-Glusberg, A., and Park, H. (2009). *2010 Census Language Program:*

Pretesting of Census 2010 Questionnaire in Five Languages. (Statistical Research Division's

Research Report Series, Survey Methodology #2009-01). Washington, DC: U.S. Census Bureau.

Pan, Y., Landreth, A., Hinsdale, M., Park, H., & Schoua-Glusberg, A. (2010). Cognitive

interviewing in non-English languages: A cross-cultural perspective. In J. Harkness, B.

Edwards, M. Braun, T. Johnson, L. Lyber, P. Mohler, B. Pennell, T. Smith (Eds.), *Survey

Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 91-113).

Berlin: Wiley Press.

Potaka, L., & Cochrane, S. (2002). *Developing Bilingual Questionnaires: Experiences from New

Zealand in the Development of the 2001 Maori Language Survey*. Paper presented at the

International Conference on Questionnaire Development, Evaluation, and Testing

Methods (QDET), Charleston, SC.

Schoua-Glusberg, A. (2006). *Eliciting education level in Spanish interviews*. Paper presented at

the American Association of Public Opinion Research Conference, Montreal, Canada.

Schoua-Glusberg, A. (1992). *Report on the Translation of the Questionnaire for the National

Treatment Improvement Evaluation Study*. Chicago: National Opinion Research Center.

Tannen, D. (2005). Interactional Sociolinguistics as a resource for Intercultural Pragmatics.

Intercultural Pragmatics, 2, 205-208.

Willis, G. (2005). *Cognitive interviewing: A tool for questionnaire design*. Thousand Oaks, CA:

Sage.

Willis, G., & Zahnd, E. (2007). Questionnaire design from a cross-cultural perspective: An

empirical investigation of Koreans and non-Koreans. *Journal of Health Care for the Poor

and Underserved, 18,* 197-217.

Willis, G., Lawrence, D., Hartman, A., Kudela, M.S., Levin, K., & Forsyth, B. (2008).

    Translation of a tobacco survey into Spanish and Asian languages: The tobacco use

    supplement to the Current Population Survey. *Nicotine & Tobacco Research, 10,* 1075-

    1084.