A case study of improving and evaluating consumer survey translation

Mandy Sha
RTI International
msha@rti.org

Jennie Lai
Nielsen[1]

Abstract

The inherent challenge of managing the translation of multilingual materials for global research is to develop a systematic translation process considering the varying translation expertise and resources available locally. In the United States, consumer survey materials are typically first designed in English (the source language) and then translated into the target language(s) with culture-specific tailoring for markets outside of the United States. In an effort to standardize the translation method globally, three experimental approaches were developed to assess the quality of translation tailoring as a function of the varying resources of expertise and tools available in the local markets. These three approaches involved different translator qualifications: a centralized professional translation service within a large company headquartered in the United States, a local professional translation service, and a company employee who is a native speaker and familiar with consumer surveys. In addition, a toolkit of translation reference materials (including translation validation form, translation input document, terminology bank and/or step-by-step translation procedures) was developed and tailored to the qualification of the translators. The experiment followed the standardized approach of Translation, Review, Adjudication, Pretesting, and Documentation (TRAPD) developed for survey research, and the translation output is evaluated based on the toolkit provided in addition to expertise and resources available for translation and review locally. The face-to-face recruitment materials of a large consumer panel survey in China were used for evaluation of the proposed translation methods from English to Chinese. For each approach, we conducted a review of translation quality based on a qualitative evaluation through a small number of one-on-one interviews in China and a quantitative evaluation of scoring the translation by assigning issue codes that were designed using sociolinguistics principles. This paper will discuss one of the three translation approaches (using a centralized professional translation service), describe the lessons learned therefrom, and share recommendations for adapting and implementing translation standards as well as areas for future research.

Keywords: translation, cross-cultural, consumer surveys

[1] This research was conducted while the co-author was with Nielsen. The co-author is currently with Google, 1600 Amphitheatre Parkway, Mountain View, CA 94043.

1. Introduction

Following a sound translation method is critical to collecting comparable survey data in all languages. When conducting consumer surveys in the global markets, we must consider tailoring translation approaches based on the expertise or resources available in the local country. To ensure consistency in the translation quality across countries, more research should be conducted to assess the quality of translation from both qualitative and quantitative standpoints.

Materials used for consumer surveys can include respondent recruitment literature, data collector training materials and job aids, and interviewer- or respondent self-administered questionnaires. In the United States, these materials are usually developed first in English and then translated into the local language(s) used by the global markets outside of the United States because there is often not sufficient time or budget for simultaneous development of the two language versions. In some markets, it may not always be possible to identify well-trained translators versed in survey research. Thus, it is important to identify a viable translation process that can be adapted and tailored to the varying level of expertise and resources available in the local markets. Following a sound translation method is critical to collecting comparable survey data in all languages. However, we are not aware of prior research that examines effect of tailoring translation approaches based on expertise or resources available in the local country.

A large consumer and media research company headquartered in the United States developed and funded an initiative to standardize and identify workable procedures for translations, with the goal of ensuring consistency of data quality across global markets. Using an English-Chinese translation project in China as a case study, this paper discusses the development of a systematic translation process based on the TRAPD (Translation, Review, Adjudication, Pretesting and Documentation) model described by Harkness (2003) for survey research. The team translation approach of TRAPD enlists a group of people who work together from the start to the end of the translation process. The translator first produces the draft of the translation to be reviewed with the assigned reviewer, and then one or more adjudicators determine whether translation can be pretested and finalized for survey deployment.

An experiment is conducted to evaluate three translation approaches by using translators and reviewers of varying levels of expertise (in response to varying resources available for each country) and providing a translation toolkit tailored to their expertise (in an effort to compensate for limited skill set or experience). The evaluation is both qualitative and quantitative. The qualitative review reports how the translation evolved from the source text to the initial translation, and to the revisions after extensive reviews and one-on-one pretest interviews. The quantitative method scores the output of the translation based on key measures. In this paper, we will examine the lessons learned for one of the three approaches conducted (using a centralized professional translation service within the research organization), share recommendations for adapting and implementing this translation standard, and discuss the research in progress for the other two approaches.

1.1 Background on the Consumer Survey Translation Project

China is one of the biggest markets in the world, and a panel survey was launched in China in 2011 to measure fast moving consumer goods (FMCG) that panellists

buy. In an effort to monitor shopper behaviour through a consumer panel, in-country data collectors conducted face-to-face recruitment of panel participants beginning in Shanghai (there is also an online component to recruitment that is not part of our project, because the materials were not ready). The shopper data is collected via handheld scanners that transmit the data back to the central office. Households across diverse Chinese provinces were contacted, and the Chinese language recruitment materials were translated from the English language materials with allowances for the urban-rural differences in China.

The face-to-face recruitment materials included six separate forms and questionnaires (some had overlapping information) that were developed in English, the source language. Considerations about designing materials for translation will be discussed in *Section 3. Designing Materials for Translation*. The forms included (1) household recruitment talking points; (2) household recruitment persuader (FAQs); (3) recruitment brochure; (4) recruitment showcard, panel expectation showcard, and confidentiality showcard; (5) household recruitment survey; and (6) household recruitment case control form and its instructional sheet.

Harkness (2010) identified team translation to be the current best practice for survey translation. Following this guidance, this study assembled a team of translators, reviewers, and adjudicators to translate these English materials into Chinese, the target language in this project. The team members were native speakers of Chinese and had prior experience with translation. Before the translation process began, they were trained with a toolkit of background information on the China Consumer Panel Project, the purpose and use of the materials to be translated, and the guidelines for translation procedures whenever possible. The toolkit for translation was tailored to translators according to their skill set and experience (it may not always be possible to hire/use professional translators, depending on environmental factors or cost limitations for each project). The process began with the translator's producing translation drafts for discussion with the reviewer and adjudicator. One or more persons can participate in each step, and the translations, review comments, and changes are documented using part of the toolkit. We tested one of the three translation approaches in 2011, and continue to progress toward completing the experiment and achieving the main goal of this research initiative, which is to identify workable processes for global markets. All three approaches are discussed in detail in *Section 2. Identifying Translation Approaches for Testing*.

To pre-test the translation provided by the translation team, a survey methodologist conducted six one-on-one Chinese language interviews with a convenience sample of company employees and non-employees. The interviews took no longer than 45 minutes and were recorded with verbal consent from the volunteers (men and women) in Shanghai, China, one of the sites for the consumer panel. None of the respondents had direct knowledge of the consumer panel study for which the translations were completed, but were reacting to the materials and translations for the first time during the interview. Three respondents were company employees who have extensive experience with consumer survey research. They were shown the *F2F Recruitment Case Control Form* and *Recruitment Talking Points and Persuader* that will be used by data collectors who share these respondents' backgrounds. The other three respondents did not have any consumer research knowledge and resembled the actual householders who will be recruited in the panel. These respondents were shown

the *Panel Expectation Showcard, Confidentiality Showcard,* and *Household Recruitment Questionnaire* that will be given to actual recruits.

As the interviewer, the survey methodologist followed an interview protocol guide, in which a language expert identified key terms and messages to focus on in the translation. She began the interview by providing a brief introduction of the purpose of the interview, and then presented the assigned materials to the interviewees. To elicit detailed responses, she used scripted probes in the protocol guide, as well as spontaneous probes. Key findings from both tests were incorporated in the final Chinese translations.

2. Identifying Translation Approaches for Testing

For this experiment, we designed three approaches. Each approach uses a team of translators, reviewers, and adjudicators – similar to what is described in the Guidelines for Best Practice in Cross-Cultural Surveys (Survey Research Center 2010) – and has an accompanying toolkit. As shown in Table 1, the qualifications for the translators and reviewers vary across approaches, while the adjudicator is a U.S.-based survey methodologist who is familiar with the intent of the materials. These three approaches mirror possible scenarios when translation projects are commissioned in an emerging market.

Table 1. The experimental translation methodologies

Approach	Translator	Reviewer	Adjudicator	Pre-test
#1	Centralized, professional translation service within the company	China-based data collection team manager	U.S.-based methodologists	Convenience sample of respondents in China
#2	Translation firm with good past performance	China-based data collection team manager	U.S.-based methodologists	Convenience sample of respondents in China
#3	China-based data collection team member	China-based survey specialist	U.S.-based methodologists	Convenience sample of respondents in China
Control	China-based data collection team member	China-based survey specialist	None	None

The three approaches are described as follows.

- Approach #1 can be considered as having the highest standard for rendering good survey translation quality. The team consisted of in-house professional translators and local field data collection staff and survey methodologists who conducted the reviews and adjudication. The initial translators, reviewers, and adjudicators complemented each other's skills in translation, survey methodology, and data collection. Specifically, the individuals who produced the initial translation in the case study were professional translators working for the company, and therefore familiar with consumer research and China.
- Approach #2 is similar to Approach #1, except that it entails using a professional translation vendor outside of the company. In the experiment, the initial translator is a professional translation vendor

located in China. Even though this vendor needs to have a strong track record, its performance may not be as good as the translator in the first approach because its degree of consumer research and familiarity with the company is unknown.

- Approach #3 seeks to reflect circumstances that are less than ideal (but nonetheless a reality of research conducted in emerging markets) – namely when professional translation staff either internally or externally are not available, and it is necessary to use nonprofessional internal resources for translation. In this experiment, the initial translator was a native speaker who worked for the local data collection team and was not a professional translator. The reviewers were survey managers in China, but the adjudicators are survey methodologists in the United States.

At the beginning of this experiment, we had designed a control condition for each approach. As the experiment progressed, we realized that the design could not be reliably implemented. We were not confident that the same individuals assigned to both the experimental and control conditions could remain unbiased. It was also challenging to manage this level of complexity remotely from the United States. In addition, for each approach, we needed to identify qualified resources from a limited pool of individuals, including strong adjudicators who spoke the local language, namely Chinese. Thus, the control condition would be defaulted to a potential (likely) scenario in the global markets, where the initial translation and its reviews were to be completed by local field data collection and management staff who were readily available.

As indicated in Pan and De la Puente (2005), toolkits are consequential because they provide the team with the necessary background information and training. Thus, part of our design is that the presence and content of the toolkit also differed across approaches. All experimental approaches are provided with a set of basic toolkits, including the *Translation Purpose and Background* (originated from the U.S. Census Bureau) that covers the source and target languages; the what, when, and how of the materials to be used; and the known linguistic issues. Another basic component is the *Translation Input-Output Spreadsheet*, which as demonstrated in the completed Approach #1, served a triple role for: documentation on the source language text; target language translation from translator, reviewer, and adjudicator; and comments provided at each stage. The Translation Input-Output Spreadsheet contains the following fields: (1) content of source language (English); (2) description (explanation or additional information provided to facilitate translation); (3) target language (Chinese translation) from translator; (4) comments from translator; (5) target language (Chinese translation) from reviewer; (6) comments from reviewer; (7) target language (Chinese translation) from adjudicator; (8) comments from adjudicator; and (9) final version of content in target language. Furthermore, the final Chinese translation, which incorporated pretesting results, was included in this document as well. It was formatted to allow full view of the translation changes from start to finish and documentation of the reason for the changes.

The toolkits provided to the teams in Approaches #2 and #3 are designed to be more complete in an effort to understand whether they could compensate for limitations in skill set or experience. Approach #2 provided a word bank to ensure consistency of language throughout the research material itself and across all the related materials. Its addition may bridge the gap of research or organizational knowledge for language usage appropriate to the research materials. In Approach

#3 (using internal resources for translation rather than a professional translation vendor), a translation guideline called "translation-by-number" would also be provided for a step-by-step instructional guide for the translation procedures and key considerations. It is understood that these reference materials cannot and should not substitute for certified translators trained for both the source and target languages. However, in the absence of professional resources, it is hypothesized that the translation toolkit can raise quality standards compared to if no reference materials provided at all. We will examine whether this hypothesis is supported once the experiment has been completed.

3. Developing the Materials for Translation

The materials used for the China consumer panel study were developed first in English with the purpose of producing a Chinese translation for fielding in China. This parallel approach for materials development was chosen to combine previously fielded materials (i.e., English language content borrowed from other surveys) and new content designed specifically for the China study.

For example, when the *Talking Points* document was developed, the lead survey methodologist considered the purpose and utility of this document: it describes each step, provides the required or optional texts that field data collectors should deliver to the households, and what recruitment materials they can refer to during a conversation. The field data collector team, which had extensive experience with consumer panel service in Europe, was able to provide feedback to enhance the operational design and wording in the Chinese context. Similarly, the paradata (i.e. data about the data) collected in the *Case Control Form* can be used to evaluate key performance indicators (KPI) for recruitment efforts. The China-based statistical sampling team provided feedback on whether the form covered all the required data and how it could be improved so that the data will be linked to sampling and reporting documents.

The materials were drafted by two Chinese-speaking survey methodologists from a U.S. office. Feedback was provided by the data collection team in Shanghai, China and their non-Chinese data collection governance team in France, as well as by the statistical sampling team in China. All of them provided comments to be incorporated. Several rounds of discussions and reviews took place before the materials were finalized, and documentation on the changes was recorded in e-mails and comment boxes in the English language materials.

The parallel development and multicultural membership of the materials developers offered several advantages. First, it allowed for Chinese cultural and linguistic considerations to be adopted before the source materials were finalized. For example, the word "panel" was removed from all English language materials even though it describes the project and the respondents' participation therein. The developers pointed out that the Chinese translation "样本框" was too technical. The phrase was hard to understand because panel studies were still a novelty in China. They decided to refer to the panel as a study or use the name of the project only. Second, because the developers were also stakeholders in the study, they were able to identify potential misunderstandings about the intent and wordings of the materials that would otherwise be overlooked. For example, the Talking Points document explains that the household was selected through "statistical sampling" (The phrase says: "Since we cannot ask every household to participate, we use statistical sampling to make sure that everyone has an equal

chance of being included in the Nielsen Family.") This phrase was changed to "scientific method" because rural households are expected to be less familiar with the concept of sampling, but would respect the authority of the sciences. In addition, because there are several consumer research organizations operating in China, the respondent screening questionnaire asks about the household's involvement in panels organized by other companies. When this question was drafted, the developers inherited a list of panel names in English and in Chinese from a survey conducted less than 3 years ago. During the translation stage, the developers were reminded that these names needed to be verified because China's changing regulations and interests as a consumer market might have affected the presence and validity of some of these panels. The translators did not necessarily have the resources to verify the Chinese language panel names because some are not publicly available.

Yet, even with this collaboration from multicultural materials developers, some problems that originated in the English language materials did not become apparent until the actual fielding phase. For example, to motivate households to participate, the company's recruitment literature often uses the phrase "your opinion counts." The developers recognized that this is a very American expression and communicated to the translator that by "counts," they mean that the household being recruited matters. The Chinese translation recommended by the translation team was: 你的看法至关重要 (your opinion is extremely important). After fielding the recruitment brochures with this phrase for a few months, the data collectors in China noted in a debriefing session that the brochures were not impactful to the Chinese respondents because they do not consider voicing an opinion to be so important. Fundamentally, they do not believe that their voice will be heard or valued. Furthermore, this particular panel requires participants to scan purchased goods rather than answer an opinion survey. It would motivate the respondents more to see an action-oriented phrase: 您的参与至关重要 (your participation is extremely important). This change was adopted.

Furthermore, during the early development phase, a consultant who studied the Chinese-speaking population in the United States raised an issue about using the word "community" in the phrase "there are millions of households. You will not only represent yourself, but also other households in your community." She noted that the concept of a community (社区) has less immediacy in the Chinese language and suggested referring to concrete geographic designators like the city or province where respondents live. This recommendation was not accepted by the materials developers because they felt that community (社区) can be understood. For this reason, the phrase using 社区 was not probed specifically for translation pretesting. However, after fielding the translated materials, the data collector team needed a way to improve this message to motivate the households. It appeared that even though the term "community" can be understood, people may encounter difficulties relating to the concept and being a representative of the larger community. As a result, the English language materials were modified and the term "community" was replaced by the geographic marker of "city," as follows: "By becoming a Nielsen Family, you will not only represent yourself, but also thousands of households in your city."

There were also some indications that changes in the source document may be warranted so that the translation could be more effective, but these could not be agreed upon. For example, messages about the company's commitment to panel participants include providing the households with newsletters that communicate

the benefits of participation, including both tangible benefits (to receive the newsletter, or to redeem gifts and participate in a lottery) and intangible benefits (representing yourself and others in your area). The benefits were described in this way: "interact closely with Nielsen and other Nielsen families."(定时给您发送简报，让您时刻了解尼尔森以及其他尼尔森家庭的最新情况). Even though the translation was accurate, two of the three pre-test respondents who were interviewed about the translation did not see the newsletters as a benefit. They also did not understand why interactions with the company relate to their participation. This may not be surprising given that using newsletters to communicate with members of a program is more common in the United States than in China. Ideally, the English language phrase would be rewritten to address these reactions. However, this is how the message has always been communicated in the company's literature, and therefore changes were not considered to maintain consistency. This is something that can be flagged for future research when developing materials to be translated in Chinese that need to convey a sense of benefit.

4. Translating the Materials from Source to Target Language

As discussed in *Section 2. Identifying Translation Approaches*, all three experimental translation approaches would use a team composed of translators, reviewers, and adjudicators as recommended in the current literature. They handled identical materials but the qualifications of the initial translator varied.

The methodologists who developed the English source language materials communicated directly with the translator through conference calls. The first communication was to give the translator an overview of the project using the *Translation Purpose and Background* document in the toolkit, and then run through the materials for translation to ensure that the intent of the materials is accurately understood by the translator. For a translation of this size (approximately 77% unique words to be translated in six documents), the initial translation took 2 weeks. For each set of translated material, the Chinese text is documented in the *Translation Input-Output Spreadsheet* in the toolkit. The Chinese translation is entered into the column marked "Target Language from Translator," while the "Comments from Translator" column is used to document any comments from the "Translator" to share with the "Reviewer" and "Adjudicator". Additional information about the material, such as definition of a term, is provided in the "Description" column.

After the initial translation, reviews and adjudication led to revisions. For all three approaches, the reviewer is a bilingual staff member with the local data collection team. Using the input-output spreadsheet, she was able to review the translation and the source (English) version side by side. This person was also involved as a developer of the English material, so she was knowledgeable about the content of the materials. Changes to the translation during the review phase are entered into the column of "Target Language from Reviewer," accompanied by justifications for the changes. One and a half weeks were allotted for this step.

Next, the *Translation Input-Output Spreadsheet* was shared with the adjudicator. The adjudicator selected between the versions produced by the original translator and the reviewer, and often provided additional changes. In this case study, the adjudicators were also part of the team that designed the source

language materials, providing them with a clear understanding of the intent and purpose of the original wording.

For example, one of the respondent screening questions contains the word "grocery," which is a commonly used term in English and in the United States to describe store-sold products like produce, meats, and food items and FMCG, such as baby speciality, personal care products, drinks, etc. The dictionary translation uses 杂货 to mean grocery in the Chinese language. However, 杂货 refers to inexpensive items at a traditional neighbourhood store in China that may be closer to a 'dollar store' in the United States. Through multiple rounds of discussion, the adjudicators, who were also part of the development team, suggested using a descriptive phrase to communicate the concept of grocery: 食品和日用品 (foods and goods). This type of discussion was important to document, because the final translation could not be found in the dictionary. Rather, it was localized to reflect what Chinese consumers perceive "grocery" to be. However, we found that it was difficult to document all the details. As much as possible, documentation about changes and comments were kept within the translation input-output spreadsheet.

At the end of this process, the translation had gone through multiple revisions and the finalized version agreed upon by the team was ready. Next, in consultation with the translation team, another methodologist developed a scripted protocol guide to pre-test selected translations with potential users in China. Because not all the translated documents could be tested in one sitting because of length, the selections focused on key terms and messages.

4.1 Qualitative and Quantitative Assessment of Translation Quality
We conducted a qualitative and a quantitative review of the translation to assemble evidence regarding improvements.

Qualitative Review of Translation Improvement: A qualitative review of how the translation evolved from the source text to the initial translation and the revisions showed that the improvements were largely Chinese language specific. The reviews corrected errors, replaced high-register terminology that may be confusing, and simplified difficult words and sentences.

Table 2 shows how the translation evolved for an instruction that reminds the data collector about preparing for a household visit.

Table 2. Error in translation of "to have"

English (source)	Chinese (target) – initial	Chinese (target) – reviewer	Chinese (target) – adjudicator
Make sure to have all the materials listed in the "Recruitment Checklist".	确保所有材料均记录在"招募清单"上 (make sure all the materials are listed in the "Recruitment Checklist")	确保包含了在"招募清单"上的所有材料 (make sure to include all the materials in the "Recruitment Checklists".)	确保带了所有包括在"招募列表"上的材料 (make sure to bring all those materials included in the "Recruitment Checklist")

In this example, the verb "have" means "to possess" and it instructs the data collector to ensure that they have all the materials that are in a checklist. However, the initial translator interpreted this verb to be auxiliary, meaning "to have listed." As a result, the translation instructs the data collector *to list* all the materials, rather than *to have* all the listed materials. This error was detected during the review stage and a correction was made. In the adjudicator stage, the team was

able to discuss and improve the translation by using the word 带了 (to bring), which is closest to the original intent of the English source text.

Table 3 provides another example to illustrate how the translation improved. In this example, the English source text is a question on the *Recruitment Persuader* document, presented in a question form in the tone of the respondent. As shown in Table 3, the initial translation used 周期 (periodicity) to ask for the period of participation. Although the translation is adequate, the word choice is of an unnecessarily high register and would not be understood by respondents with lower educational levels. The review phase simplified the translation by asking 多久 (how much time), which was agreed upon by the adjudication.

Table 3. High-register terminology choice and difficult words

English (source)	Chinese (target) – initial	Chinese (target) – reviewer	Chinese (target) – adjudicator
How long do I need to be Nielsen Family?	我参加这个项目的周期有多长？ (How long must I participate in this project?)	我需要参加这个项目多久？ (How much time do I need to participate in this project?)	我需要参加这个项目多久？ (How much time do I need to participate in this project?)

Even though all the translations were correct, and had been greatly improved by the team of translators, reviewers and adjudicators, our cognitive pretesting effort revealed issues that came to light when the communicative effect of the translation was discussed with respondents. The utility of cognitive testing is consistent with prior literature describing the methodology for cognitive testing of multilingual survey translations (Sha & Pan, 2013) and the testing of Chinese survey translation (Sha et al., 2012). Table 4 provides an example of how a correctly translated phrase improved its delivery of the intended message.

Table 4. The translation delivers the intended message

English	**Your home was specially selected.** Since we cannot ask every household to participate, we use a scientific method to make sure that every household has equal chance to be included in Nielsen Family.
Translation prior to testing	我们特别选择了您的家庭参与本次项目。 (We in particular selected your family to participate in this project.)
Revision post-testing	您的家庭被选中参与本次项目是一个特别的机会。(The selection of your family's participation is a special opportunity.)

As seen in Table 4, the English source text intends to impress upon the household the specialness of their participation by explaining that the selection is random and not every household could participate. The initial translation provided by the translator for the phrase "your home was specially selected" was not modified by the reviewers or the adjudicators because it was correctly translated. However, when this phrase was pretested, respondents pointed out that the message of the translation achieves quite the opposite of what was intended. Instead of being impressed by the specialness of the selection, the respondents felt targeted. All respondents agreed that by revising the translation to say that the special yet random selection is a 特别的机会 (special opportunity) alleviates the concern.

Some improvements in the translation were made after the translation was fielded with actual panel respondents. In the questionnaire used to recruit households, the field interviewers collect the address of the household by first asking for the street name. It was correctly translated to say 街道名称 and our pre-test respondents did not exhibit any problems with the translation. A portion of the actual respondents, however, experienced difficulties. They thought the translation asked about the name of their neighbourhood association, which is called "街道居委会". It is because the term 街道 for "street" could also refer to a neighbourhood in general. As a result, the translation was shortened to 街名/路名 and no confusions were reported.

Quantitative Review of Translation Improvement: To examine whether the translation improved from the initial version, two methodologists reviewed and coded translation issues at three progressive stages of the translation process: (1) translator to reviewer stage (the initial translation was reviewed and modified by the reviewers); (2) reviewer to adjudicator (the reviewers' modified translation was agreed upon by the adjudicators); and (3) adjudicator to the final version (the final version incorporated feedback from the translation pretesting and the additional quality control to standardize vocabulary used across documents).

The methodologists coded the translation issues in the *Talking Points* document. (This document was not selected for analysis because we did not have the resources to code all of the translated documents.) Using a coding scheme similar to that developed by Pan and Fond (2014), we coded issues using four categories: Linguistic Rules, Cultural Norms, Social Practices, and Production Errors. The Pan and Fond coding scheme is guided by sociolinguistic approaches to language and culture to evaluate translated questionnaires by classifying translation issues. (Pan and Fond coding also includes a code category called "User Error" because the codes were developed to summarize cognitive interview data with actual respondents. This category does not apply to our project.) However, as soon as we began the coding exercise, we realized that this project needed additional categories to highlight some of the issues produced by the different translation approaches under test. Table 5 shows a complete list of codes used, a description of what they are, and the number of occurrences in each stage of the translation.

As seen in Table 5, *translation errors* and *production errors* were only found at the translator-to-reviewer stage, where the initial translation was reviewed for the first time. This result was not surprising. One would expect that mistakes introduced by the translator or in the copyediting process would be caught early in the initial translation review stage. It also shows the value of having reviewers for QC purposes. We did not observe any issues related to "social practice", but that is probably because the particular document we used to conduct the exercise did not have major equivalency problems.

When looking at the total number of translation issues identified by all three stages, one would expect more to be found during an earlier stage of the process, such as when the initial translation was reviewed for the first time. Yet, the majority or 60% of the issues were observed at the reviewer-to-adjudicator stage. This was when the translation had been reviewed previously, and the adjudicators modified it for further improvement. Examining the specifics, most modifications made by the adjudicators were in the *linguistic rules* and *standardization* categories. For example, the number of issues identified because of *linguistic rules* tripled during the reviewer-to-adjudicator stage as compared with the

previous stage (24 vs. 7). These are language-specific issues: problems associated with grammar, sentence structure, and wording. Looking at the specifics, the majority of the changes by the adjudicators improved upon the translation issues that the reviewers had detected during the earlier stage. Some, however, were never detected.

Why do our findings show more translation issues identified later in the process? This is probably because of a disparity in the qualifications between the reviewers and the adjudicators. In this project, the adjudicators were methodologists with extensive experience in survey research. The reviewer was a bilingual professional who is familiar with field work in China, but not a methodologist. Such disparity in qualifications was a factor we noted in earlier sections; it can be fundamentally challenging to locate language and survey experts in emerging markets. In addition, almost all standardization issues were identified later, not earlier (9 vs. 1). In this project, it was only at the adjudication stage that the translation team decided to formally standardize some of the terminology across the documents for translation (e.g. always use 条码指南 to mean 'barcode booklet for the scan'). Therefore, it was not surprising that standardization issues were chiefly revealed during a later stage. Evidently, none of these issues would have been resolved if the initial translation had been reviewed by native speakers only, without survey methodologists, or if it had not been reviewed at all. For other translation approaches, we plan to incorporate a word bank that standardizes the word choice of the phrases used. If these translation approaches do not result in standardization problems, then this would support the theoretical assumption that skills can be compensated with the toolkit.

Table 5. Coding of translation issues

Codes	Description	Translator to Reviewer	Reviewer to Adjudicator	Adjudicator to Final Version	Total
1. Translator errors	Wrong translation (e.g., "dog" translated to "cat") or translator left blank and unsure how to translate	2	0	0	2
2. Linguistic rules	Problems with words, word order, sentence structures	7	24	1	32
3. Cultural norms	Same concept, but different ways of expressing it	0	0	3	3
4. Social Practice	No equivalence, or different concepts or practice, problem associated with respondents' unfamiliarity with survey conventions	0	0	0	0
5. Production errors	Copying/pasting errors and problems associated with document editing and formatting	1	0	0	1
6. Standardization issues	Did not use pre-established standardized translation	1	9	0	10
	Total	11	25	4	40

The coding results from the final stage, adjudicators-to-final version, also suggested that the translation showed a significant improvement. This was when the translation was finalized after the adjudicators had modified the translation,

and selected translation items were pretested with potential users of the translation in one-on-one interviews. (Some changes were also made in the final version to ensure consistency across documents.) Fewest issues (four instances) were observed at this stage, and most were *cultural norms* that referred to the cultural frame of reference, such as ways of doing things and styles of communication, and not confined to the word and sentence level. This result was in the direction that we expected because, during the one-on-one interviews, the respondents provided their reaction to the translation based on how they interpreted the message. During the translator-reviewer-adjudicator process, no such communication could occur. As a result, these issues had evaded the translation reviews but surfaced during the one-on-one interviews with Chinese speakers. Conversely, more issues were sourced to *linguistic rules* during the translator-reviewer-adjudicator process because those reviews concentrated on language-specific issues.

5. Discussion and Areas for Future Research

This case study has provided useful insight about developing translation standards for consumer surveys. First, as we expected, identifying the appropriate language experts may be the greatest challenge for translation projects in global markets. We had to modify the scope of our experiment simply because it was not realistic for us to identify multiple sets of language experts. However, this does not mean that translation quality should be compromised. Early identification of a translation vendor and establishing a long-term relationship with it may mitigate the resource limitation.

We also learned that the materials development team is crucial to the success of the translation. In our project, because the majority of the materials developers were familiar with the Chinese language and culture, when the source language materials were finalized they were already linguistically and culturally tailored to some degree. Several potential problems with the intent and wordings of the materials were identified before translation began. In addition, the field-data collection team members became more comfortable with the content because they had participated in the development process.

Even though a parallel development process and a multicultural development team offered many advantages, not all issues could be identified while the materials were being developed. Having an experienced development team may maximize the chances of identifying potential problems in the translation. If schedule and budget allow, it would also be helpful to have an "advance translation" – that is, a rough rendering during the English source-material development stage to detect possible translation problems and revise the source material accordingly (Harkness, 2003). With this approach, issues like "grocery" could have been resolved during the development phase. Logical flaws such as pointing to random yet special selection for participation could also have been highlighted and revised in the source text. Conducting advance translation is feasible when there is a multicultural, multilingual team of experienced materials developers. Furthermore, parallel development is also time-consuming. It requires considerable effort to document and integrate feedback in multiple rounds of revisions. However, we found that the benefit of this approach was significant because the source items were not necessarily inflexible, and if one was found to be problematic for Chinese rendering, it was usually changed.

Upon examination of our experiment's first approach (using a centralized professional translation service within the research organization), the translator-reviewer-adjudicator pretesting process was shown to improve the translation. The qualitative assessment suggested that the reviews conducted by the reviewers and adjudicators corrected language-specific issues, while the one-on-one interviews uncovered issues beyond words and sentences, such as ensuring that a linguistically correct translation also delivers the intended message. When the issues identified in the translation revisions were quantified through a coding exercise, we found that the reviews conducted for the initial translation were able to catch errors from the translator and the copyediting process. Reviews from the later phases detected additional issues related to improving word choice and sentence structures, including some that had not been identified before. This was probably because the adjudicators were survey methodologists who were familiar with both the intent of the source text and how the translation would be used for data collection. This suggests the utility of a team approach, enabling members with differing levels of language or survey expertise (as we would expect in emerging markets) to complement each other's skills.

The quantitative results corroborated the qualitative assessment. The pre-test, performed via one-on-one interviews with Chinese speakers, uncovered more issues related to *cultural norms,* while the reviews largely revealed issues concerning *linguistic rules*. When our experiment is complete, the coding results will allow us to compare the instances of *translation errors* across the different initial translations, and provide evidence of the impact of translator qualifications on translation quality. Incorporating a quantitative review of translation quality may be used in quality assessments by managers who do not speak the target language.

Certainly, quantitative coding and a team approach to translation can be more costly and time-consuming. Coders need to be trained to accurately identify and capture different issues: assessing translation quality is not as simple as determining "correct versus incorrect," but also whether it conveys the intended message while using easy-to-understand words and syntax. For the translation, we had to conduct several rounds of reviews to ensure that the Chinese-language materials reflected the English original. Keeping the documentation up to date on decisions and reasons for changes was sometimes challenging. If translation is factored into the project planning phase, then its cost and scheduling may be balanced against competing priorities and the need for quality. On several occasions we also had to make the decision to revise the translation after it was fielded. As discussed in the previous section, the field interviewers found that a common Chinese term for "street" has evolved to mean "neighbourhood" when used in a certain context. Allowing time in the schedule and budget for pre-field testing, and possibly revising the materials and translation after fielding, will improve the delivery of the intended messages.

Our next step is to complete the experiment and provide recommendations on whether and how to use any of the three approaches that were designed here to measure the quality of translation by using translators and reviewers of varying levels of expertise, together with a toolkit tailored to their expertise. For example, we will implement a word bank of standardized terminology in the translation process, and examine how much improvement this brings to the quality associated with issues of linguistic rules, cultural norms, and social custom, even when one cannot locate a translator with strong qualifications: we at least expect that this may minimize production errors and translator errors as defined in our coding

scheme. The completion of testing will also allow us to discuss limitations associated with implementing the different translation approaches, and evaluate their respective impacts on translation quality.

Acknowledgement

References

Pan, Y., & De la Puente, M. (2005). Census Bureau guideline for the translation of data collection instruments and supporting materials: Documentation on how the Guideline was developed. (Survey Methodology #2005-06).U.S. Census Bureau. Retrieved from http://www.census.gov/srd/papers/pdf/rsm2005-06.pdf

Pan, Y., & Fond, M. (2014). Evaluating multilingual questionnaires: A sociolinguistic perspective. *Survey Research Methods, 8*(3), 181-194.

Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. Van de Vijver, & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp.35-56). Hoboken, NJ: Wiley.

Harkness, J. A., Van de Vijver, F. J. R., & Johnson, T. P. (2003). Questionnaire design in comparative research. In J. A. Harkness, F. J. R. Van de Vijver, & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp.19-34). Hoboken, NJ: Wiley.

Harkness, J. A., Villar A., & Edwards, B. (2010). Translation, adaptation, and design. In J.A. Harkness, M. Braun, B. Edwards, T.P. Johnson, L.E. Lyberg, P.Ph. Mohler, B-E. Pennell & T.W. Smith (Eds.) *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 117-140). Hoboken, NJ: Wiley and Sons, Inc.

Sha, M., Park, H., & Pan, Y. (2012). *Translation review and cognitive testing of American Community Survey (ACS) language assistance guides in multiple languages*. Prepared for the American Community Survey Library Collections. Retrieved from http://www.census.gov/acs/www/Downloads/library/2012/2012_Sha_01.pdf

Sha, M., & Pan, Y. (2013). Adapting and improving methods to manage cognitive pretesting of multilingual survey instruments. *Survey Practice, 6*(4), 1–8. Retrieved from http://surveypractice.org/index.php/SurveyPractice/article/view/253/pdf

Survey Research Center (2010). *Guidelines for best practice in cross-cultural surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from http://www.ccsg.isr.umich.edu/