Not a Copy and Paste: Makings of Non-English Data Collection Instruments

FedCASIC 2019 Workshop

Lily Kapaku, Brianda Perez
Content, Language, and Data Products Branch
Decennial Census Management Division, U.S. Census Bureau, Washington, DC
Multilingual Automated Data Collection Instruments

Purpose

• The 2020 Census counts every person living in the United States, once, only once, and in the right place.
• Help reduce language barriers non-English speakers may encounter.
• Optimize the non-English content across data collection modes and operations.
• Provide translations that are accurate, meaningful and culturally relevant.
# Multilingual Automated Data Collection Instruments

## Background

<table>
<thead>
<tr>
<th>Instrument/Material</th>
<th>2020 Non-English Support</th>
<th>2010 Non-English Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Self-Response</td>
<td>12 Non-English Languages</td>
<td>Not in Scope</td>
</tr>
<tr>
<td></td>
<td>- Spanish, Chinese (Simplified), Vietnamese, Korean, Russian, Arabic, Tagalog, Polish, French, Haitian Creole, Portuguese, Japanese</td>
<td><em>Fulfillment and Be Counted Questionnaires in 5 non-English languages, available via phone request or pickup: Spanish, Chinese (Simplified), Korean, Vietnamese, Russian</em></td>
</tr>
<tr>
<td></td>
<td>Language Selection Criteria: 60,000 or more housing units</td>
<td><em>Language Selection Criteria: 100,000 or more housing units</em></td>
</tr>
<tr>
<td>Census Questionnaire Assistance (Telephone) Interviewer Based</td>
<td>12 Non-English Languages</td>
<td>5 Non-English Languages</td>
</tr>
<tr>
<td></td>
<td>- Spanish, Chinese (Mandarin and Cantonese), Vietnamese, Korean, Russian, Arabic, Tagalog, Polish, French, Haitian Creole, Portuguese, Japanese</td>
<td>- Spanish, Chinese (Mandarin and Cantonese), Korean, Vietnamese, Russian</td>
</tr>
<tr>
<td></td>
<td>Telecommunication Device for the Deaf also available</td>
<td>Telecommunication Device for the Deaf also available</td>
</tr>
<tr>
<td></td>
<td>Language Selection Criteria: 60,000 or more housing units</td>
<td>Language Selection Criteria: 100,000 or more housing units</td>
</tr>
<tr>
<td>Enumerator Instrument (Mobile) Interviewer Based</td>
<td>Spanish (hand-held)</td>
<td>Not in scope</td>
</tr>
</tbody>
</table>
Multilingual Automated Data Collection Instruments Process

Translate
- Translation
- Review
- Quality Assurance

Develop
- Programming
- Formatting
- Functionality

Pretest
- Pre-Testing: Cognitive and Usability Testing

Update
- Adjudication
- Final Translation Updates
- Quality Assurance

Finalize
- Instrument Updates
- Usability and User Acceptance Testing
Challenges and Findings

Style, Terminology, and Other Considerations
Challenges and Findings: Translations
Consider Multiple Meanings of English Terms

- English content with context specific translations

Translation of Vietnamese (race) vs. Vietnamese (language)
Challenges and Findings: Translations

**English Terms That Should Be Included in Translations**

- Some English terms are included because there is no equivalent term or because the English term is the optimal term.
- These English terms are emphasized using parentheses and/or italics

- In a nursing home or group home
- In a jail or prison
- En un hogar de ancianos o *nursing home*, o en un hogar de grupo
- En una cárcel o prisión
Challenges and Findings: Translations
Consider Whether Responses Are Needed in English

• Some responses can be translated as necessary, others cannot
• Instruction to provide names in English added to address output requirements
Challenges and Findings: Translations
Translations Developed In Both Past And Future Tense

• Census Day is April 1, 2020. Respondents can start responding in March 2020.
• Verb tense changes dynamically based on when the respondent is completing the census.

Portuguese translation:
"will be living" = "estará morando"
"were living" = "estava morando"
Challenges and Findings: Translations
Mode-Specific Formality Considerations

- For interviewer instruments (telephone and mobile), mode-specific translations are crucial.
  - Need to incorporate politeness markers.

Spanish incorporates polite terms for you ("usted") and your ("su") instead of informal terms

"How is SAM related to JIMMY? SAM is JIMMY's..."
Example shows님(Korean honorific) after Name fills.

SAM님은 JIMMY님과 어떤 관계 입니까?
SAM님은 JIMMY님의...
Challenges and Findings: Translations

Cannot Repurpose Translations

- Translations need to have 1-1 relationship with English text and be context specific.
- For example, cannot use a translation found in another question for “Other”
  - Другим родственником/родственницей is the translation of “Other” found on another screen, but is the equivalent of "Another relative/relative"
Challenges and Findings: Translation and Design

Emphasizing Non-English Text

• There are instances where English text is emphasized in questionnaires in all capital letters.
  • Not all languages have capitals or emphasize text using capitals.
  • Substitute other appropriate styles, such as underlining, to ensure the emphasis is not lost.
Challenges and Findings: Translation and Design

Date of Birth

• When providing the date of birth, respondents may use different date formats (MM/DD/YYYY vs DD/MM/YYYY).

• Use numbers AND month names

<table>
<thead>
<tr>
<th>Haitian Creole</th>
<th>Polish</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 - Janvye</td>
<td>01 - styczeń</td>
<td>01 - يناير</td>
</tr>
<tr>
<td>02 - Fenvye</td>
<td>02 - luty</td>
<td>02 - فبراير</td>
</tr>
<tr>
<td>03 - Mas</td>
<td>03 - marzec</td>
<td>03 - مارس</td>
</tr>
<tr>
<td>04 - Avril</td>
<td>04 - kwiecień</td>
<td>04 - أبريل</td>
</tr>
<tr>
<td>05 - Me</td>
<td>05 - maj</td>
<td>05 - مايو</td>
</tr>
<tr>
<td>06 - Jen</td>
<td>06 - czerwiec</td>
<td>06 - يونيو</td>
</tr>
<tr>
<td>07 - Jiyè</td>
<td>07 - lipiec</td>
<td>07 - يوليو</td>
</tr>
<tr>
<td>08 - Out</td>
<td>08 - sierpień</td>
<td>08 - أغسطس</td>
</tr>
<tr>
<td>09 - Septam</td>
<td>09 - wrzesień</td>
<td>09 - سبتمبر</td>
</tr>
<tr>
<td>10 - Oktôb</td>
<td>10 - październik</td>
<td>10 - أكتوبر</td>
</tr>
<tr>
<td>11 - Novenm</td>
<td>11 - listopad</td>
<td>11 - نوفمبر</td>
</tr>
<tr>
<td>12 - Desanm</td>
<td>12 - grudzień</td>
<td>12 - ديسمبر</td>
</tr>
</tbody>
</table>
Challenges and Findings: Design
Response Option Design

- Use of X, can be interpreted as "No" or "Does not apply"
- Instead, use fill-in, checkmark or circle for affirmative response.

Radio button

Check box
Challenges and Findings: Design
Optimize How Translations Appear

- Text expansion causes translations to appear differently than the English content.
- Translations need additional work to be optimized for non-English data collection instruments.

Name translations without html code

Russian internet instrument with name field translations (with and without <br> code)
Challenges and Findings: Design
Optimizing Non-English Buttons

• Length of translation should also be assessed for optimal design
  • Example: vary font and size in buttons

• Directional elements should incorporate correct directions
  • Arrows for languages read right-to-left, such as Arabic: right for “Previous” and left for “Next”
Challenges and Findings: Design
Optimize Field Order and Layout

- Though Arabic is read right to left, phone numbers are usually read left to right.
- Need to correctly place 4-digit field.

Respondents entering (123) 456-7890 might insert (789) 456-123X
Best Practices Summary

**Translations:**
- Use professional translators and follow industry best practices for quality assurance.
  - Use plain language and provide translator notes to ensure English meaning is clear.
  - Include English terms in the translation when needed for clarity.
  - Consider how the translation sounds in spoken modes, and incorporate appropriate formal terms.
  - Consider dynamic content needs (fills, singular/plural, verb tense adjustments)
- Do not "repurpose" translations throughout an instrument. Each translation must have a context-specific relationship with the English text.
- Use usability and cognitive testing to identify critical issues.

**Design**
- Consider how the translation looks (directional text, size, line breaks, formatting)
- Consider whether visual elements might be misunderstood (example: marking responses with an X).
Questions?
Contact Information

Lily Kapaku
Content, Language, and Data Products Branch
Decennial Census Management Division
U.S. Census Bureau
301-763-0189
lily.kapaku@census.gov

Brianda Perez
Content, Language, and Data Products Branch
Decennial Census Management Division
U.S. Census Bureau
301-763-5536
Brianda.Perez@census.gov
2020 Census
Connect with Us

Sign up for and manage alerts at
https://public.govdelivery.com/accounts/USCENSUS/subscriber/new

More information on the 2020 Census:
http://www.census.gov/2020Census

facebook.com/uscensusbureau
twitter.com/uscensusbureau
youtube.com/user/uscensusbureau
instagram.com/uscensusbureau
pinterest.com/uscensusbureau