SERISS Experiment for Testing Close versus More Adaptive Questionnaire Translation Approaches: First Results from the Translation Sessions and from the Field (Work in progress)

Brita Dorer, GESIS (Mannheim – Germany)
Lydia Repke, GESIS (Mannheim – Germany)
Ana Villar (London)

CSDI, Limerick, 26 March 2018
Overview

• Background: close versus adaptive translation
• Background: Research goals / research design
• Translations & first findings (qualitative)
• Measurement approach (quantitative)
1. Background

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654221.
Dilemma: close vs. adaptive translation

• Most surveys: “Ask-the-same-question (ASQ)"
• No definition of what ‘same question’ means
• Researchers make different interpretations
  – ‘As close as possible’ at the word level
  – Others focus on making stimulus functionally equivalent and believe this requires adaptation
• No empirical evidence yet on which interpretation leads to more comparable data
Deciding level of closeness: example

How much of the electricity used in [country] should be generated from each energy source?

Energy sources: Coal, Natural gas, Hydroelectric power, Nuclear power, Solar power, Wind power, Biomass

- A very large amount
- A large amount
- A medium amount
- A small amount
- None at all
Deciding level of closeness: example (cont'd)

• Advance translation and translation verification suggested in most languages 'amount'
  → share / portion / part
• BUT portion posed limitations on the answers: you can't have a very large portion of all energy sources
• TEP: Natural language use vs. safe translation?
• Which version yields most comparable data?
2. Research goals

/ Research design
Research goals / questions

• What to do when "dilemmas" come up?
• Which approach is better: 'as close as possible' vs. 'encouraging adaptation'
  – Better = yielding more comparable data
• Under which conditions is that approach better? Does the effect differ by...
  – Language: are some languages more resilient to close translations than others?
  – Type of adaptation: factual, linguistic...
  – Degree of adaptation: are some adaptations "too much"?
Overall model

Independent variable: Experimental manipulation
2 translation approaches

Dependent variable: Measurement quality
Measurement comparability

Moderating variables:
1 Distance between
   1a. Language (linguistic structure)
   1b. Country/culture (different realities)
2 Type of adaptation
Research design

• 2 sets of instructions produced: close vs adapted
• 2 languages: Slovene (SL) & Estonian (ET)
• For each language 3 teams and 3 sets of items:
  – 2 parallel Translations following experimental instructions
  – Review / Adjudication meeting with instructions present
• Manipulation checks
  – Audio taping
  – Follow up interviews with translators
• Data collection: Cross-National Online Survey (CRONOS) panel
  – 60 items
Close = 'As close as possible'
Adapt = 'Encouraging adaptation'

<table>
<thead>
<tr>
<th>Items</th>
<th>Team 1</th>
<th>Team 2</th>
<th>Team 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-20</td>
<td>Close</td>
<td>Adapt</td>
<td>Adapt</td>
</tr>
<tr>
<td>21-40</td>
<td>Adapt</td>
<td>Close</td>
<td>Adapt</td>
</tr>
<tr>
<td>41-60</td>
<td>Close</td>
<td>Adapt</td>
<td>Close</td>
</tr>
</tbody>
</table>
3. Insights from the translations
Translation sessions: Estonian (ET) and Slovene (SL)

- Took place in summer 2017
- Team approach (2 translators + 1 reviewer/adjudicator)
- All reviewers experienced in questionnaire translation
- Almost all translators experienced in questionnaire translation
- None of them from ESS, ISSP, SHARE translation teams
- Each team 3 batches
- At least 1 week between the 3 batches
- 2 instructions documents (close – adaptive)
- All Review sessions recorded (audio): Manipulation check
Findings from translation sessions

• Follow-up interviews with all 6 reviewers (3 ET and 3 SL)
• Follow-up interviews with translators of 3 last teams (2 SL teams + 1 ET team)

- Not everybody realised the difference between the approaches! (1 ET, 1 SL team)

- Most common approach: usually all experienced reviewers translate as close as possible to source (comparability) but adapt where necessary
Ongoing: qualitative aspects

- Native speakers in ET and SL
- Analyse
  a) translation templates / final translations
  b) review sessions (audio recordings)
- WHAT was actually asked, what were the differences
- Combine with qualitative results
4. Measurement approaches / insights from the field
Analytical strategy

• Dependent on the items
• Possible methods
  – Multigroup Confirmatory Factor Analysis (MGCFA)
  – Item response theory
  – Web probing
  – Questions from benchmark surveys
  – Content analysis of the review meetings: also learn about how instructions worked
## Sample

<table>
<thead>
<tr>
<th>Country/exp. group</th>
<th>EE</th>
<th>GB</th>
<th>SI</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>173</td>
<td>209</td>
<td>194</td>
<td>576</td>
</tr>
<tr>
<td>2</td>
<td>153</td>
<td>223</td>
<td>217</td>
<td>593</td>
</tr>
<tr>
<td>3</td>
<td>173</td>
<td>201</td>
<td>204</td>
<td>578</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>499</td>
<td>633</td>
<td>615</td>
<td>1.747</td>
</tr>
</tbody>
</table>

*Note.* In GB the three experimental groups are the same with respect to the questionnaire.
Example – National Identity

• 2013 ISSP Module on National Identity
• Dimensions measured by 5 items
  – Nationalism (2 items, 5-point scale)
  – Constructive patriotism (3 items, 4-point scale)
• Method: MGCFA (Davidov, 2009; Meitinger, 2017)
• 7 groups to test for measurement equivalence
  – EE (3 groups)
  – GB (3 “groups” ≡ 1 group)
  – SI (3 groups)
Example – National Identity

World would be a better place if people were more like in own country

Own country is better than most other countries

Proud in the way democracy works in own country

Proud in own country’s social security system

Proud in own country’s fair and equal treatment of all groups in society
Example – National Identity

- Scalar invariance not supported by previous findings (Davidov, 2009)
- Possible problems with terms
  - E.g.: pride, social security system, democracy, all groups of society
  - Pride with how social security system works has been shown to load negatively on nationalism in GB (Davidov, 2009)
Next steps - Discussion

• Carry out / complete quantitative and qualitative analyses

• Discussion: How to test for measurement equivalence for single items?
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654221.

THANKS!

Brita, Lydia and Ana