TREND MEASUREMENT IN INTERNATIONAL ASSESSMENT SURVEYS

3MC International Conference 2016
Chicago, USA, July 25-29 2016
Based in Brussels, Belgium and Philadelphia PA, USA

Language services for multilingual, and multinational tests, assessments, and surveys

A network of 220 linguists in over 75 countries

Active membership ITC, ESRA, WAPOR, AAPOR, ATP, EMA, CSDI
OECD: PISA, TALIS, AHELO, PIAAC, E&S ONLINE

UNESCO: LAMP, WEI, LMP (with ACER)

EU-OSHA: ESENER-2 (TA, Adjudication, Documentation)

EU-FRA: MIDIS II (TA, Adjudication, Documentation)

IEA: PIRLS, TIMSS, TEDS-M, ICCS, ICILS

World Bank: STEP, AES

EU: ESS, SHARE, INVEDUC
A. Source Optimization

A1 controlled writing
A2 file preparation: parsing, segmentation, locking untranslatable content
A3 Translatability Assessment
A4 create project-specific rules

B. Preparation of Translation & Adaptation Process

B1 create glossaries
B2 create style guides
B3 create language-specific rules
B4 translation & adaptation notes
B5 trend mgt (content transfers)
B6 monitoring tool, documentation
B7 train translators (& verifiers)

C. Translation and Adaptation Process

C1 double or single translation
C2 reconciliation
C3 (team) adjudication
C4 consultation (domain experts)
C5 proofreading

D. Linguistic Quality Control Process

D1 automated checks
D2 translation verification
D3 errata and updates mgt
D4 review of verification feedback
D5 post-verif final checks
D6 reports, updated TMs, post mortem
Surveys aim at collecting data: knowledge, skills, competences, background information, etc.

If periodical data collections are planned, it is usually of interest to also measure change over time.

So items that have been administered in the past are administered again (\(=\) trend items), usually in conjunction with newly developed items.
if you want to measure change, don’t change the measure

Exceptions, however, might be necessary:

- Questions can become outdated.
- Errors undetected due to poor item functioning
- Change in survey or test delivery mode
Following comment by Japanese national team, item developers decided the trend item could probably (*) be “saved” with an edit in stimulus (wording of item unchanged).
(*) Subject to confirmation from FT data
Example

Ruling on currency change request: okay given that the items in the test units concerned do not refer to currency/amounts mentioned in the stimulus text.

<table>
<thead>
<tr>
<th>ENGLISH SOURCE VERSION</th>
<th>ITEM-SPECIFIC TRANSLATION/ADAPTATION GUIDES AND JUSTIFICATION (IN ENGLISH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>£</td>
<td></td>
</tr>
<tr>
<td>€</td>
<td></td>
</tr>
</tbody>
</table>

Latvia will adopt the Euro in January 2014 which indeed would require adapting euro amounts into euros in this unit (and also in unit shirts). This means that this unit should be entirely reviewed accordingly. It would be recommended to use the amounts as in FRA source.
Example

PISA approach: generally accepted to change spelling in trend items when a country has enacted a spelling reform.

<table>
<thead>
<tr>
<th>ITEM-SPECIFIC TRANSLATION/ADAPTATION GUIDELINE</th>
<th>REQUEST FOR CHANGE VERSUS ARCHIVE VERSION: Description of desired change (in target language) and justification (in English)</th>
<th>TRANSLATION REFEREE AND/OR TEST DEVELOPERS’ COMMENTS</th>
<th>VERIFIER INTERVENTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptation to new Portuguese orthography</td>
<td>Replace optimismo by optimismo</td>
<td>OK (change for new spelling rules agreed in PISA 2012 for trends)</td>
<td>OK</td>
</tr>
</tbody>
</table>
Theory vs. Practice

- Or, as always, someone, somewhere, think they know better and want to make a change, or a new team takes over the survey.

- Whatever the reasons, design strict procedures to filter and control changes in trend content, so that even the tiniest edit is clearly documented and its effect can be tracked.
PISA approach: for ‘outright errors’: if item was dodgy in previous cycle, it is normally corrected in current cycle. If it was not dodgy, decided on a case-by-case basis. Typos are generally corrected.

<table>
<thead>
<tr>
<th>ENGLISH SOURCE VERSION</th>
<th>ITEM-SPECIFIC TRANSLATION/ADAPTATION GUIDELINE</th>
<th>REQUEST FOR CHANGE VERSUS ARCHIVE VERSION: Description of desired change (in target language) and justification (in English)</th>
<th>TRANSLATION REFEREE AND/OR TEST DEVELOPERS’ COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Stationery, print and promotion products, technology supplies. See the general terms of the offer.</em></td>
<td></td>
<td>Last word in Lithuanian version contains spelling error, letter &quot;é&quot; missing. Should be: taisykiės</td>
<td>Ok to correct spelling mistake</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example

- Recent case: DNK asked "We have decided to change the sentence"
- "click on the NEXT arrow"
- From
- “klik på NÆSTE-pilen…”
- To
- “klik på pilen “NÆSTE”…”
- as it seems more idiomatic.”

PISA approach for ‘preferential changes’: generally rejected. After consulting the linguist and the referee, it was decided to reject this change. It’s true that it is more idiomatic, but the “old” version is not per se incorrect.
IEA Studies: a different approach

IEA approach
Linguists are asked to identify the differences in trends without expressing their opinion or indicating the type of difference:

<table>
<thead>
<tr>
<th>2011 version: line break</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016 version: no line break</td>
</tr>
<tr>
<td>2011: full stop at the end</td>
</tr>
<tr>
<td>2016: no full stop at the end</td>
</tr>
<tr>
<td>2011: nel caso che avesse avuto bisogno di qualcuno che badasse a lei.</td>
</tr>
<tr>
<td>2016: nel caso avesse avuto bisogno di qualcuno che badasse a lei</td>
</tr>
</tbody>
</table>

We’re not sure about the adjudication process (perhaps at the level of countries), but anyway “even the tiniest edit is clearly documented and its effect can be tracked”
In an ideal world, requests for changes to trend content should always be supported by data.

If an item shows differential item functioning or bias, there is a good reason to scrutinize the wording or cultural adaptations carefully and possibly to propose alternative wording to remedy that situation.

Or, one could argue that once an item seems to have worked well, even correcting a residual error is an unnecessary risk.

But it might be difficult to convince a national team that a clearly identified error should preferably remain uncorrected and be kept for the next administration.”
Under what circumstances can a test or survey question be regarded as obsolete?

It would depend on the situation: a scientific fact, or a spelling changes in country would merit a change, currency change would affect the value and therefore the numbers involved.
Excerpt from PISA2015 Verification Report:
The understanding of the “trend” procedure and the reasons for the “no changes between cycles” policy varied considerably across countries.

- Some countries understood the process and requested either no changes at all or a limited number of justified changes (i) to correct outright errors; (ii) as a reaction to an item bias detected in previous cycles (e.g. Finland, Colombia).

- Other countries requested a large number of changes that seemed mostly preferential (Country X, Country Y). Such requests typically originated from countries where the national team had changed since the previous cycle, as was the case with Country X.

- For a small group of countries, the trend materials were reviewed perfunctorily, as obvious errors were overlooked in items that had been dodgy in the previous cycle (e.g. Country Z).
Open questions

- What are the risks involved when trend material is ‘opened’ for review?

**Note on ‘centralized’ trend management (e.g. PISA2015):**
Countries do not have editing access to their trend materials (test units and questionnaires); they have the right to review these materials and make requests for changes, which are then negotiated. Agreed changes are implemented by the international project team, not by the countries (considerably less risk than when countries edit trend materials, which are then submitted for verification)
Open questions

- Is it sensible to transfer known errors across survey cycles?

Although difficult to get across as an idea, this would be a necessary by-product of the “strictest” possible approach to trend management, consisting of “No changes whatsoever to trend items, under no circumstances”. With such an approach, trend items would not be opened for review at all (cf. previous slide).
Open questions

Who should be assigned with the role of determining whether a change is acceptable to make or not?

PISA approach: the “Translation Referee” advises countries on translation plans, reviews all verification feedback and negotiates with countries on crucial issues until corrective action is agreed – liaising with item developers as needed.
Open questions

- Does the shift to a computer-based or online environment offer a new paradigm for content management over time, i.e. to what extent can documentation of changes to trend items be automated?
Open questions

- In the context of large-scale multilingual, multinational and multicultural assessments where national teams in participating countries are responsible for translation and adaptation (decentralized management), is it possible to organize the management of trend content by a single country-independent organization (centralized management) with a view to controlling the urge to revise materials?

See previous slide on ‘centralized’ trend management in PISA2015:
- Considerably less risk than when countries are allowed to edit trend materials (even if they are then submitted for verification)
- Considerably more work for the international project team
Conclusion

- Very thorough documentation of any changes made to trend items -- this way results can be analysed.

- A "centralized" trend process is not easy, and tends to be costly. However, it helps immensely in protecting trend materials.
THANK YOU!

laura.wayrynen@capstaninc.us
andrea.ferrari@capstan.be