

# Six cost-effective modules for improving master questionnaire prior to translation and adaptation

Musab Hayatli & Manuel Souto Pico  
**cApStAn™ Linguistic Quality Control**

**Second International Conference on Survey Methods in Multinational,  
Multiregional and Multicultural Contexts– Chicago, 27 July 2016**

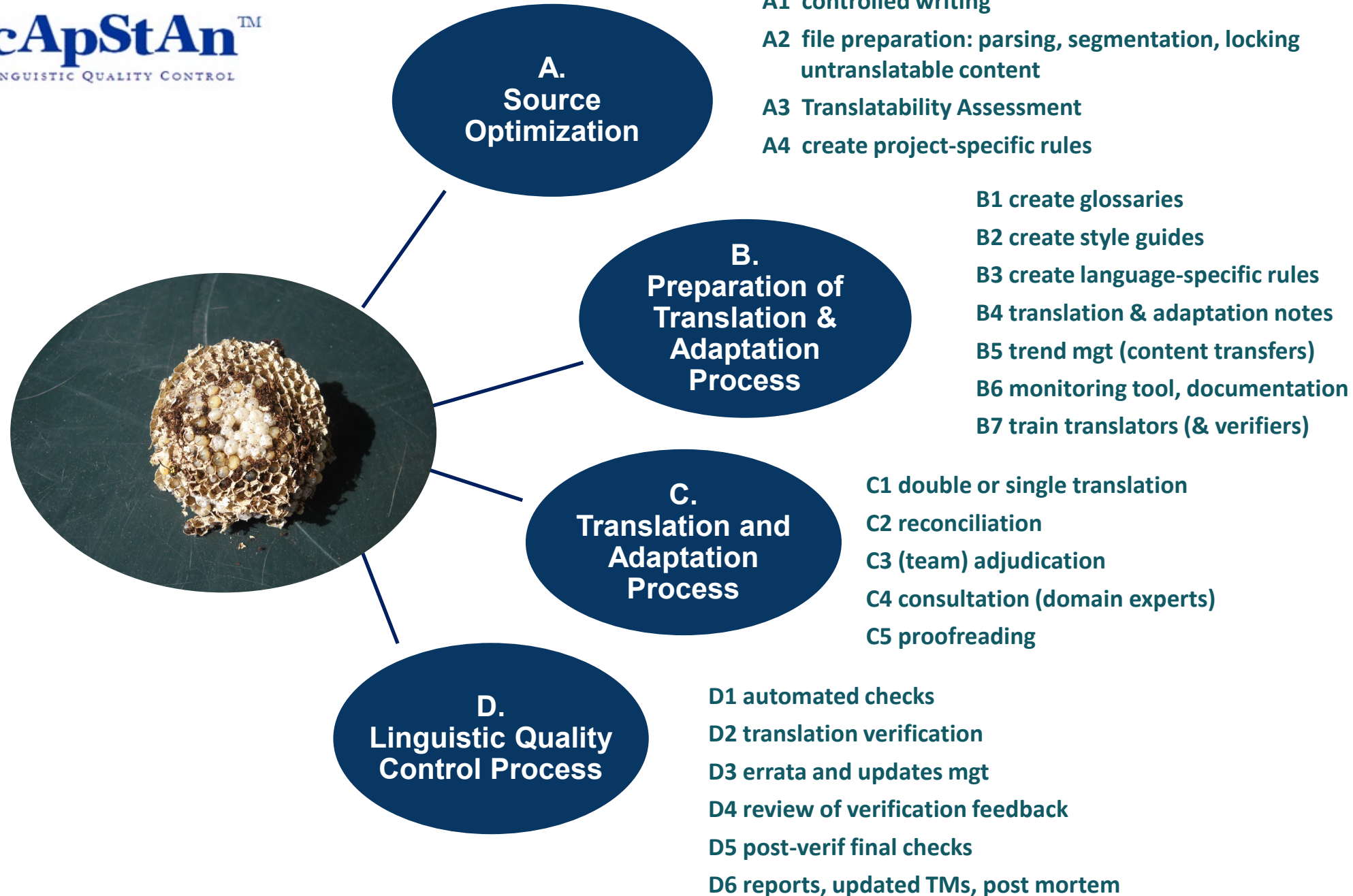
**cApStAn™**  
LINGUISTIC QUALITY CONTROL

# cApStAn<sup>TM</sup>

- 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, CSDI

# cApStAn<sup>TM</sup>

- 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**

**A2 file preparation: parsing, segmentation,  
locking untranslatable content**

**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**

## A2. File Preparation: Extraction and Parsing

- Review master version with the knowledge that text will be extracted (= separated from layout)
  - decide what is translatable and what is not
  - protect non-translatable parts (e.g. items not displayed to respondent, placeholders that stand for graphical items)
- Optimal result
  - only translatable parts of text are editable for the linguist
  - untranslatable parts are either hidden or locked (not editable)

## A2. File Preparation: Segmentation

- split text into smaller, manageable parts, called “segments”
  - segmentation points should match sentence boundaries, not style markers
  - Avoid scattering sentences across several segments  
e.g. because one word is underlined in the sentence
  - Exception: sentences that need to be completed by selecting options in drop-down menus

## A4. Create Project-specific Rules

Compliance with these rules can be checked automatically

- Rules that can be linked to attributes of the source version
- regardless which/how many target languages
  - Define max. number of digits in captions
  - Define <forced adaptations>
  - Identify the need for dynamic text
  - Determine how tags and fills must be treated





## B1. Create (and maintain) glossaries

- Creating glossaries is best practice in (pre-)translation, but remains underrated and underused
- bilingual or multilingual glossaries rather than explanatory glosses in the source language
- Creating glossaries > some work involved, but immense quality gains

# Glossaries may contain

- terms that need to be translated consistently (technical concepts, cultural entities, ...)
- fictitious names
- terms or expressions that can be translated in more than one way
- recurring prompts
- acronyms
- names of months and days
- names of questionnaire sections or modules
- names of buttons, tables and other UI elements

Adherence to the glossary can be checked automatically



## B2. Create Style Guides

- Create a generic template with all style items we want specified (for all languages in the project)
- Each linguist (language lead) adapts the template to his/her specific language variant
  - Typographic conventions (quotes, indents,...)
  - forms of address, gender-neutrality
  - notation of numbers, mathematical units
  - Currencies, time

Country: ISRAEL	Language: ARABIC	Example																						
Comment/Example	<b>ara-ISR</b> Insert your version in this column	Insert your examples in this column																						
Language variant	Which language variant are you using?	Non relevant																						
Form of address in Cognitive units (15 years old students)	Formal/Informal?	Formal																						
Numerals	European/Arabic Indic numerals <table border="1"> <tr> <td>European (descended from the West Arabic)</td> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td> </tr> <tr> <td>Arabic-Indic</td> <td>٠</td><td>١</td><td>٢</td><td>٣</td><td>٤</td><td>٥</td><td>٦</td><td>٧</td><td>٨</td><td>٩</td> </tr> </table>	European (descended from the West Arabic)	0	1	2	3	4	5	6	7	8	9	Arabic-Indic	٠	١	٢	٣	٤	٥	٦	٧	٨	٩	European numerals 1,338,613,000
European (descended from the West Arabic)	0	1	2	3	4	5	6	7	8	9														
Arabic-Indic	٠	١	٢	٣	٤	٥	٦	٧	٨	٩														
Date format	Separator, order of elements, etc. e.g. dd/mm/yyyy or dd-Month	DD/month																						
	24	جمال، الذي اشترى طابعة ب 700 شيكل بتاريخ 14 كانون الثاني.																						
	Dot	جولة لمدة 60 دقيقة تبدأ في الساعة 14:00، 15:00، 16:00.																						
	Comma, yes it is.	231.6																						
	Comma, yes it is.	1,338,613,000																						
	within punctuation signs	العتاب المخذرة لا تكمل الدورات																						
	"ABC"	يعني، لما تقول "محلي" الإثني يعتمد على وين إنت موجود.																						
Dialogues	quotation marks	الفقرة قبل الأخيرة تقول: " وجهت هذه الدراسة ضربة تالية لسمعة قصص المصممين المشهورين																						
System	Metric	قطر عجلات كرسي جاد الكبيرة يساوي 59 سنتيمتر.																						
Temperature	Celsius Symbol to the right of the figure?	14° / 28°																						
	To the right	إضافة الكمية الكاملة من السائل +2																						
	Full word, and symbol (translation)	قطر عجلات كرسي جاد الكبيرة يساوي 59 سنتيمتر.																						
	To the left	قام سيالستاتي بابتلاع أنبوب أسطواني ختبيي طوله 2 سم وقطره 0.5 سم.																						
	New Israeli shekels. To the left	أنبوب أسطواني ختبيي طوله 2 سم																						
	To the right	جمال، الذي اشترى طابعة ب 700 شيكل بتاريخ 14 كانون الثاني.																						
	To the right	دمرت الرياح القوية 40% من الأشجار متساقطة الأوراق في 60% من الأشجار دائمة الخضرة في																						

Adherence to the style guide can be checked automatically



## B3. Create Language-specific rules

- Each question mark in the master version should be a question mark in the target version.
- In Greek, this should be a semi-colon

Compliance with  
the rules can  
be checked  
automatically



## B4. Translation & Adaptation Notes

- “man-made”  
=> needs to be rendered as “manufactured by human being” (1. gender-neutral; 2. not: “caused by mankind”)
- *[as a result of this note, “man made” was changed to “human-made” in the master questionnaire]*

## B4. Translation & Adaptation Notes

- concepts, objectives, possible ambiguities or intended meaning of terms or expressions in the master version:
- explain clearly and concisely

=> the impact on quality is **immense**

- Also: specify which <adaptations> (intentional deviations from the master are mandatory, desirable, or ruled out

# CONCLUSION

- A sophisticated LQA design can be time- and work-intensive
- HOWEVER, the more (purposeful) upstream work we do, the less headaches further downstream
- [Ask professionals to] take the time to analyze (and optimize) the master questionnaire
- [Ask professionals to] segment and parse, prepare glossaries, style guides, project-specific and language-specific rules
- and to prepare concise, precise item-per-item guidelines  
=> better cross-linguistic (cross-cultural) comparability



THANK YOU



[musab.hayatli@capstaninc.us](mailto:musab.hayatli@capstaninc.us)  
[manuel.souto@capstan.be](mailto:manuel.souto@capstan.be)