




APRIL 4-8, 2019
Fairmont Royal York, Toronto, Canada

**TRANSLATION, TEST ADAPTATION
AND LINGUISTIC QUALITY
ASSURANCE PROCEDURES
IN PISA AND PIAAC**

Steve Dept
steve.dept@capstan.be



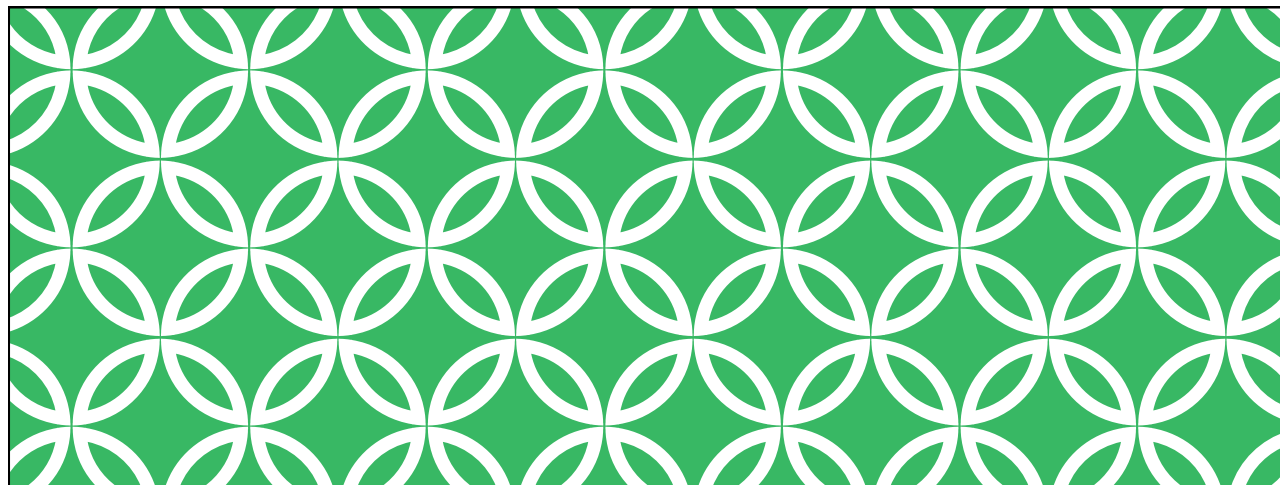
CONTENTS

SETTING THE STAGE: brief history of translation and Linguistic Quality Assurance-Linguistic Quality Control in PISA and PIAAC

LQA-LQC METHODOLOGY: procedure, tools, documentation

SHIFT IN FOCUS: More LQA, less LQC

CONCLUSIONS: What next?



SETTING THE STAGE

LANGUAGE QUALITY
MANAGEMENT IN ILSAs

SOME MILESTONES



In the late 60s: test translation changes test difficulty to the extent that comparisons across language groups may have limited validity

In the 70s: linguistic quality control methods are introduced, e.g. back translation (Brislin, 1970, 1976, 1986)

New insights are gained in how and why different forms of adaptation (to local context and usage) affect measurement

A good summary of initial breakthroughs: Adapting Achievement Tests into Multiple Languages ... (Hambleton, 2002)



FIRST COMPARATIVE SURVEYS WITH LQC

IALS (1994-1998) – adults 16-65 from 22 countries
“data that were comparable across cultures and languages”

TIMSS (1995) – 500,000 students from 3 grades and 45 countries
“rigorous procedures have been developed for the direct and inverse translation of the items into the different languages of the participating countries, in order to ensure the levels of difficulty are maintained, over and above the specific language used for the test”



THEORY OF TEST TRANSLATION ERROR (SOLANO-FLORES ET AL)

Solano-Flores, G., Contreras-Niño, L. A., & Backhoff, E. (2005).
The Mexican translation of TIMSS-95: Test translation lessons from
a post-mortem study. Paper presented at the meeting of the
National Council on Measurement in Education. Montreal, Quebec,
Canada

Conceptually very satisfying

Resource-intensive

Post hoc validation

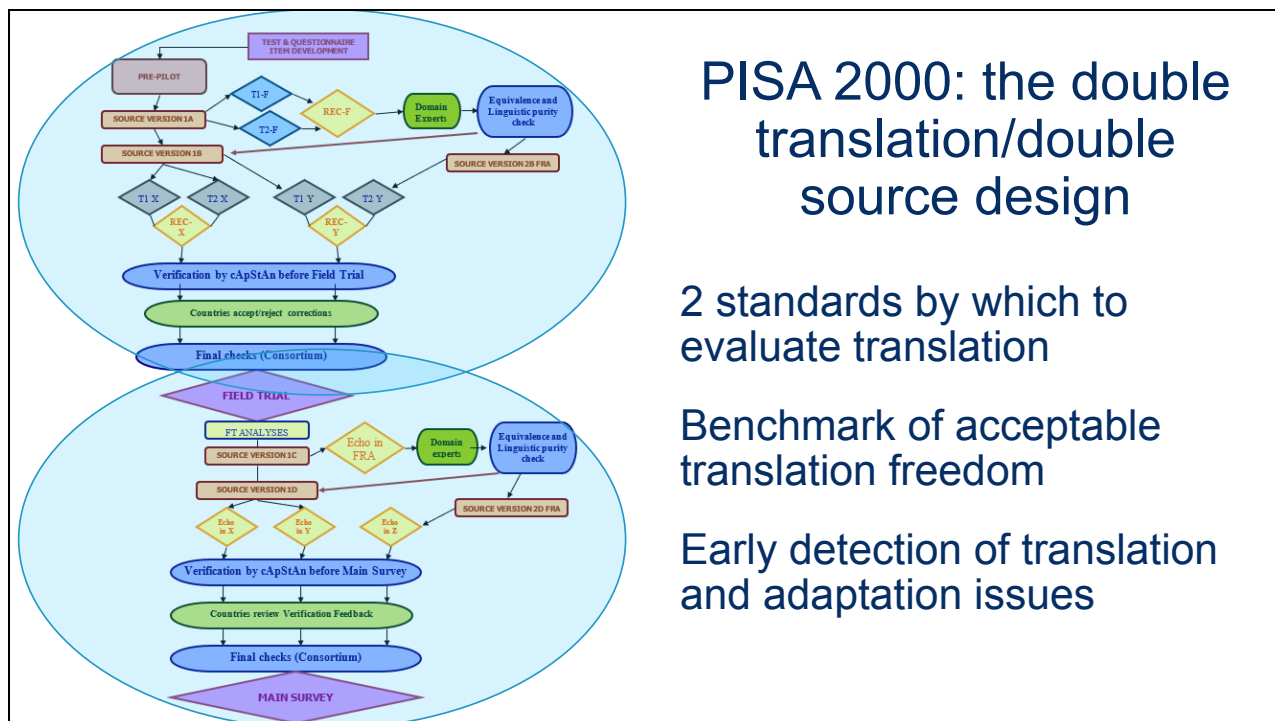


PISA 2000

Aletta GRISAY: experience in surveys, in linguistics, and in working with data : a bridge between IRT and language experts

Reference to ITC guidelines (1997 version)

PISA Translation and Adaptation Guidelines: collaborative effort, external validation.



PISA 2000: the double translation/double source design

2 standards by which to evaluate translation

Benchmark of acceptable translation freedom

Early detection of translation and adaptation issues



DECENTRALIZED TRANSLATION CENTRALIZED VERIFICATION

PISA National Centers implement double translation + reconciliation

External verification by international pool of verifiers

In PISA 2000, 28 language professionals with:

- Teaching experience (school setting)
- Translation experience (from ENG+FRA)

1999: FT verification on hard copies – no tools

PROOFREADERS' MARKS USED FOR THE VERIFICATION PROCEDURE			
	delete; take these words out		set in <u>capitals</u> (CAPITALS)
	close up; print as one word		set in <u>lowercase</u> (lowercase)
	delete and close up		set in <u>italic</u> (<i>italic</i>)
	insert; aspace		set in <u>roman</u> (roman)
	insert here		set in <u>boldface</u> (boldface)
	space evenly where indicated		hyphen
	let marked text stand as set		en dash (1965-72)
	transpose, change order		em — or long — dash
	used to separate two or more marks; often used as a concluding stroke at the end of an insertion		superscript (V as in π^2)
	set farther to the left		subscript (A as in H ₂ O)
	set farther to the right		comma
	set æ or fl as ligature: æ		apostrophe
	straighten alignment		period
	straighten or align		semicolon
	begin a new paragraph		colon
	spell out (set 5 lbs. as five pounds)		quotation marks
			parentheses
			brackets

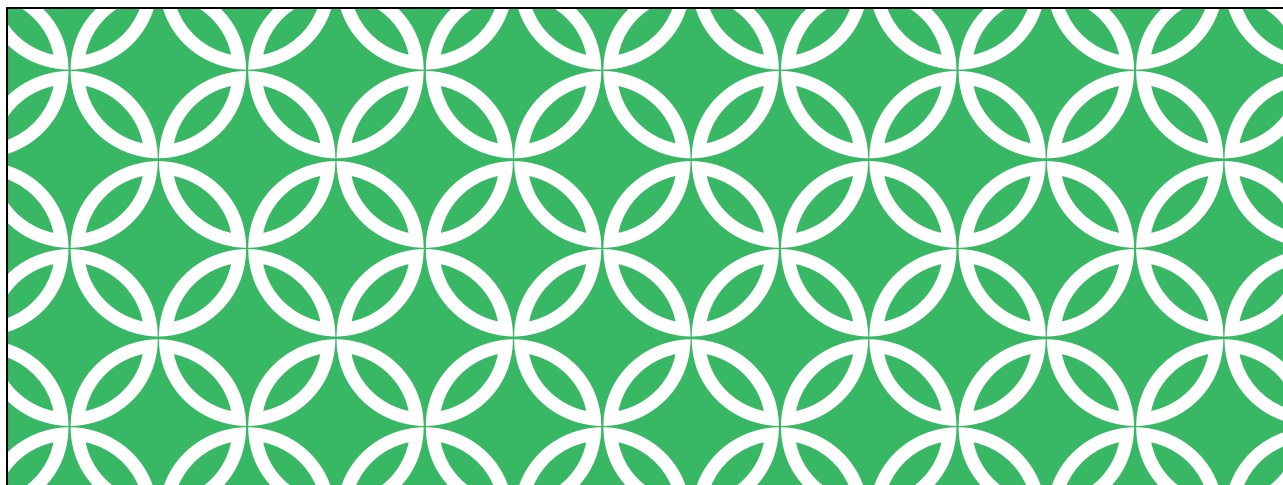
VERIFIER INTERVENTION CATEGORIES

In PISA 2006 FT, verifiers commented on issues identified, with special focus on potential equivalence issues

5,380 verifier comments, covering 42 national versions in 36 languages for 38 countries, were analysed and described with key words:

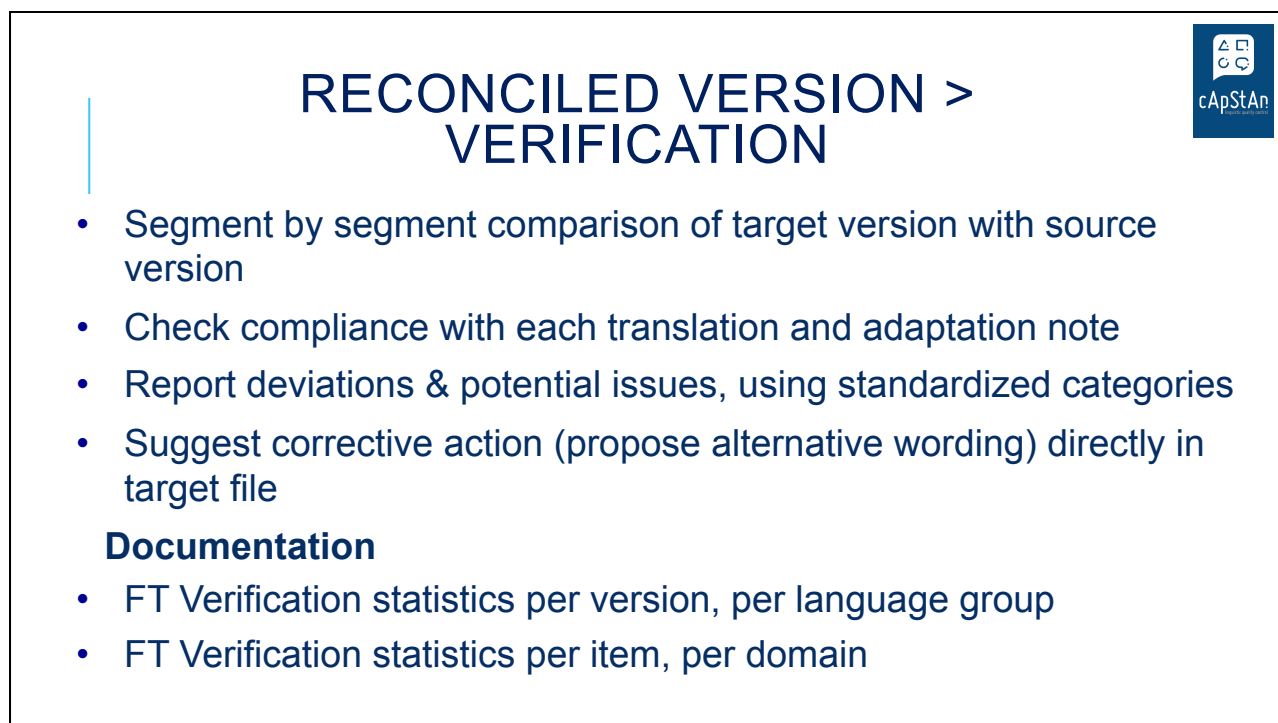
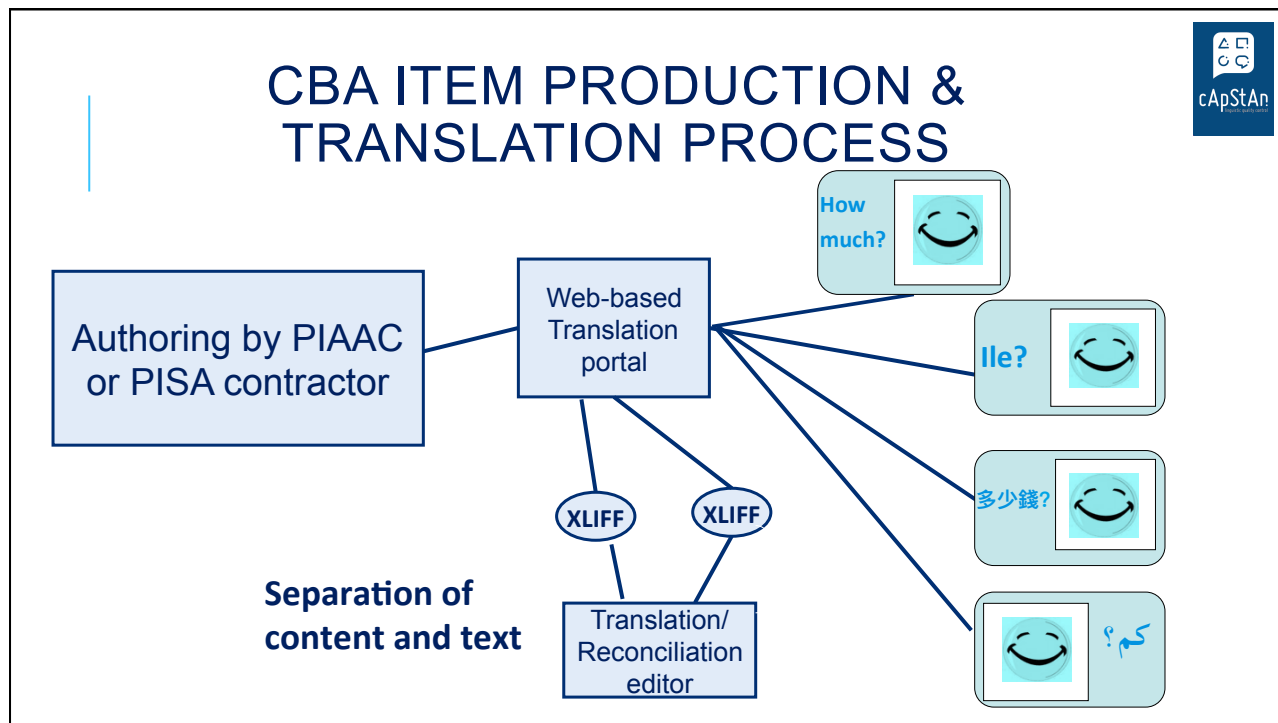
A taxonomy of verifier intervention categories was developed:

OK	No intervention is needed. The verifier has checked and confirms that the text element or segment is equivalent to source, linguistically correct, and – if applicable – that it conforms to an explicit translation/adaptation guideline. This category may also be used to report an appropriate but undocumented adaptation.
ADDED INFORMATION	An information is present in the target version but not in the source version, e.g. an explanation between brackets of a preceding word.
MISSING INFORMATION	An information is present in the source version but omitted in the target version.
MATCHES AND PATTERNS	1) A literal match (repetition of the same word or phrase) or a synonymous match (use of a synonym or paraphrase) in the source version is not reflected in the target version. Most important: literal or synonymous matches between stimulus and item and between a question stem and response categories. 2) A pattern in multiple choice items is not reflected in the target version (e.g. all but one option start with the same word, proportional length of responses options.)
INCONSISTENCY	A recurring element across units (e.g. an instruction or prompt) is inconsistently translated, and this appears to be unintentional.
ADAPTATION ISSUE	An adaptation is an intentional deviation from the source version made for cultural reasons or to conform to local usage. An adaptation issue occurs when an adaptation would be needed but was not made, or when an inappropriate or unnecessary adaptation was made.
REGISTER / WORDING ISSUE	1) Register: difference in level of terminology (scientific term >< familiar term) or level of language (formal >< casual, standard >< idiomatic) in target versus source. 2) Wording: inappropriate or less than optimal choice of vocabulary or wording in target to fluently convey the same information as in the source. This category is used typically for vague or inaccurate or not quite fluent translations.
GRAMMAR / SYNTAX ISSUE	1. Grammar: grammar mistake that could affect comprehension or equivalence, e.g. wrong subject-verb agreement, wrong case (inflected languages), wrong verb form. 2. Syntax: syntax-related deviation from the source, e.g. a long (source) sentence is split into two (target) sentences or two (source) sentences are merged into a single (target) one; or another syntactic problem due e.g. to overly literal translation of the source.
MISTRANSLATION	A wrong translation, which seriously alters the meaning. A mistranslation should always be reported with a back-translation. Note: >vague or inaccurate translation should rather be classified as a Register/Wording issue (or sometimes a Grammar/Syntax issue). This category covers cases where the source has been misunderstood, but also copy/paste errors that unintentionally result in a wrong text element or segment.
GUIDELINE NOT FOLLOWED	An explicit translation/adaptation guideline for a given text element or segment was overlooked or was not addressed in a satisfactory way.
LEFT IN SOURCE LANGUAGE	A text element or segment that should have been translated was left in source language.
MINOR LINGUISTIC DEFECT	Typo or other linguistic defect (spelling, grammar, capitalization, punctuation, etc.) that does not significantly affect comprehension or equivalence. Correcting such errors is usually not controversial and can be made in track changes without documenting them.
ERRATUM/UPDATE MISSED	An erratum or update notice has been overlooked.
LAYOUT / FORMAT ISSUE	A deviation or defect in layout or formatting: disposition of text and graphics, item labels, question numbering, styles (boldface, underlining, italics, UPPERCASE), legibility of captions, tables, number formatting (decimal separators, "five" versus "5"), etc. In computer-based materials, this includes truncated words in the preview, undesired scrolling, etc.



LQA-LQC METHODOLOGY

Procedures, tools, documentation





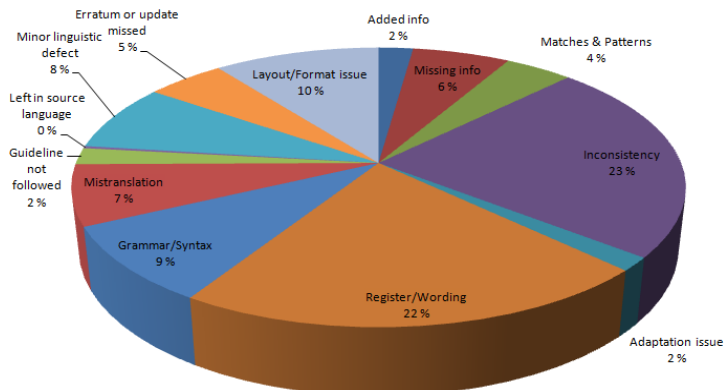
PISA2015FT		TEST ADAPTATION SPREADSHEET								
Country:	UAE	C5835 SAVE THE FISH								
Language:	Arabic									
LOCATION	SEGMENT ID	ENGLISH SOURCE VERSION	ITEM-SPECIFIC TRANSLATION/ADAPTATION GUIDELINE	COUNTRY COMMENT (ADAPTATIONS, DOUBTS)	VERIFIER INTERVENTION	VERIFIER COMMENT	CONSORTIUM REFEREE COMMENT	CORRECTION STATUS	COUNTRY POST-VERIFIER COMMENT	FINAL CHECK
Directions	02	Refer to ... on the right. Use drag and drop to answer the question.	Answer instructions: References should be to underlying in stimulus	بنفسه انفسه... الخ	Not checked	not consistent answer other comments: changed by ref	Pls see... as to keep the original content... (2015/15) Thanks		ok, just small changes to be consistent	
Text										
Table	30-53	Herbivores Primary consumer Secondary consumer Producers	Use the correct terms to feed in e.g. Biology textbooks "Producer", in the context of biology, means "an organism, especially a green plant, that builds up its own tissues from simple inorganic compounds"; this translation should be consistent with what C5607 Biol04 Arab Curriculum	القائمان... الخ القائمان... الخ القائمان... الخ	OK					
	2				Register/Wording	of data to me... of the results to... that example?	Please keep the original correction	REQUIRES FOLLOW-UP		OK
		Greek								

- OK
- Added info
- Missing info
- Matches & Patterns
- Inconsistency
- Adaptation issue
- Register/Wording
- Grammar/Syntax
- Mistranslation
- Guideline not followed
- Left in source language
- Minor linguistic defect
- Erratum or Update missed

REQUIRES FOLLOW-UP

Excel Adaptation Forms as Repositories for documentation

TYOLOGY OF ISSUES REPORTED



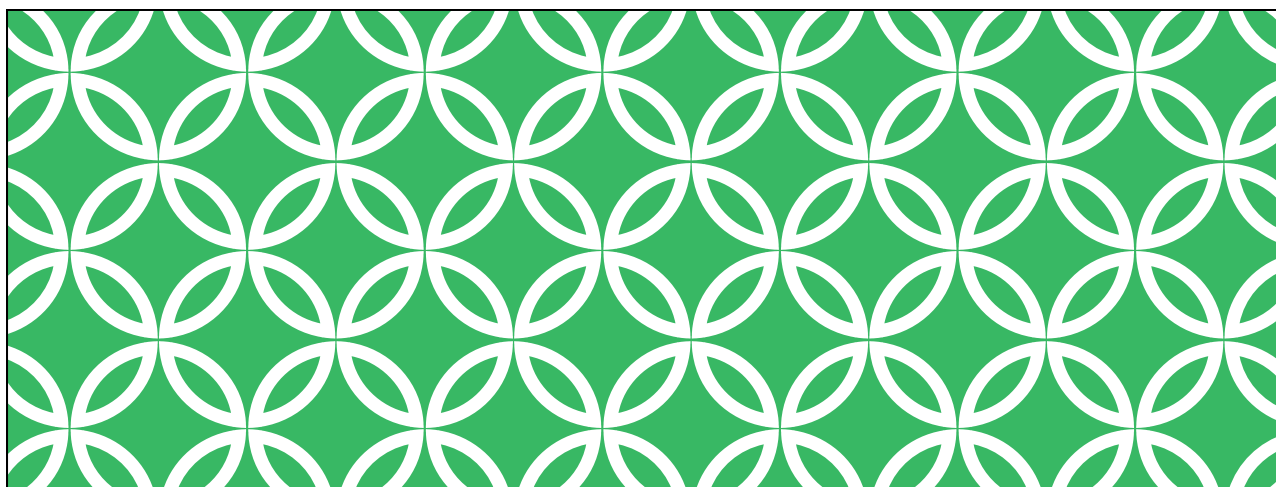
Per national version, per unit, per item, per domain, per language, per language group



VERIFIER INTERVENTION CATEGORIES PER TEST UNIT

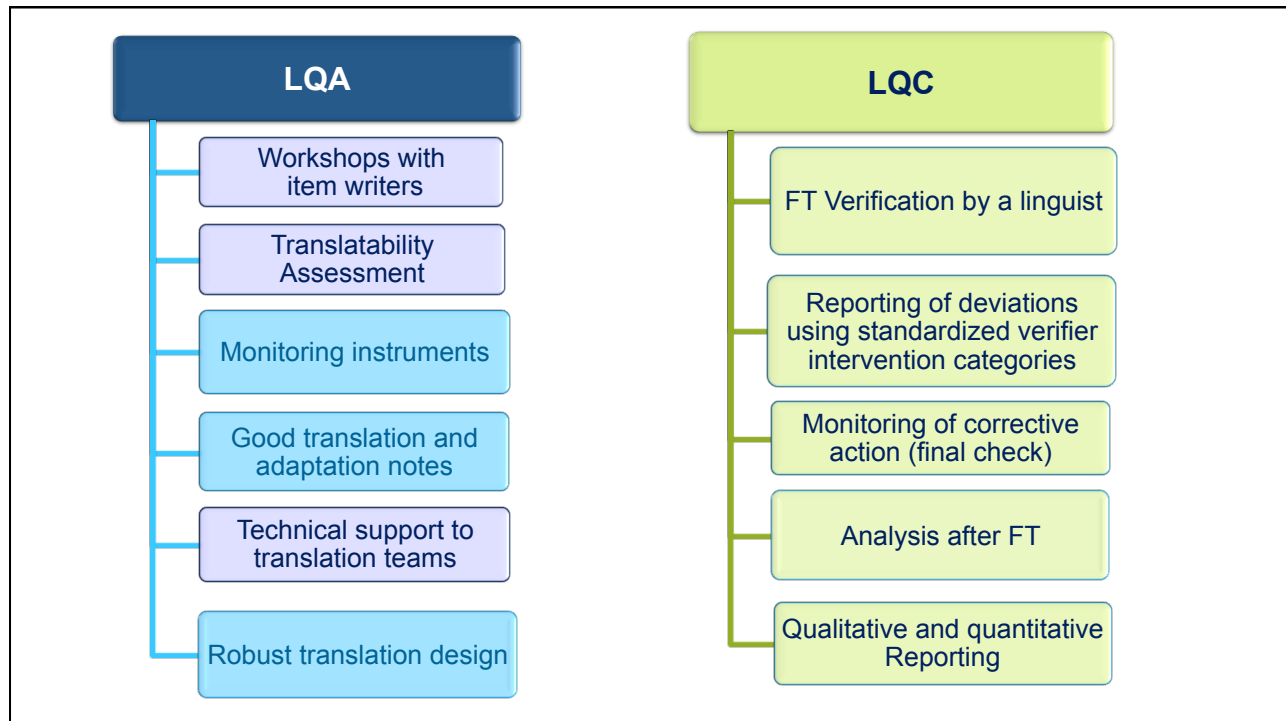
OVERVIEW OF VERIFICATION INTERVENTIONS PER CPS UNIT

Unit Code	Added info	Missing info	Matches & Patterns	Inconsistency	Adaptation issue	Register/Wording	Grammar/Syntax	Mistranslation	Guideline not followed	Left in source language	Minor linguistic defect	Erratum or update missed	Layout/Format issue	TOT CORRECTIONS	REQUIRES FOLLOW-UP	WORD COUNT
C100-Xandar	4	25	8	79	7	205	97	30	15	8	67	4	44	593	71	1988
C101-The Visit	44	60	59	267	120	690	237	100	31	7	204	28	60	1907	262	8216
C102-Field Trip	12	37	40	161	19	231	95	26	16	5	75	45	38	800	122	3540
C103-Presentation	14	43	36	148	54	298	180	53	23	15	121	57	20	1062	181	4941
C104-Meeting in the park	9	31	33	124	12	232	123	30	9	9	82	15	64	773	104	4191
C105-The Garden	30	70	99	266	39	645	263	88	34	10	177	48	27	1796	288	8194
C106-Making a film	13	31	75	217	22	403	152	48	13	4	108	1	42	1129	151	5829
TOTAL	126	297	350	1262	273	2704	1147	375	141	58	834	198	295	8060	1179	36899



SHIFT IN FOCUS

Does more LQA mean less LQC?



DRIVING FACTORS



- The transition from pencil-and-paper tests to computer-delivered assessments;
- The realization that a number of recurring translation/adaptation issues could be prevented by allocating more resources to crafting a more translatable source.
- The need to improve management of trend content (link units)
- The ambition to harness the power of translation technology



CHALLENGES AND TENSIONS

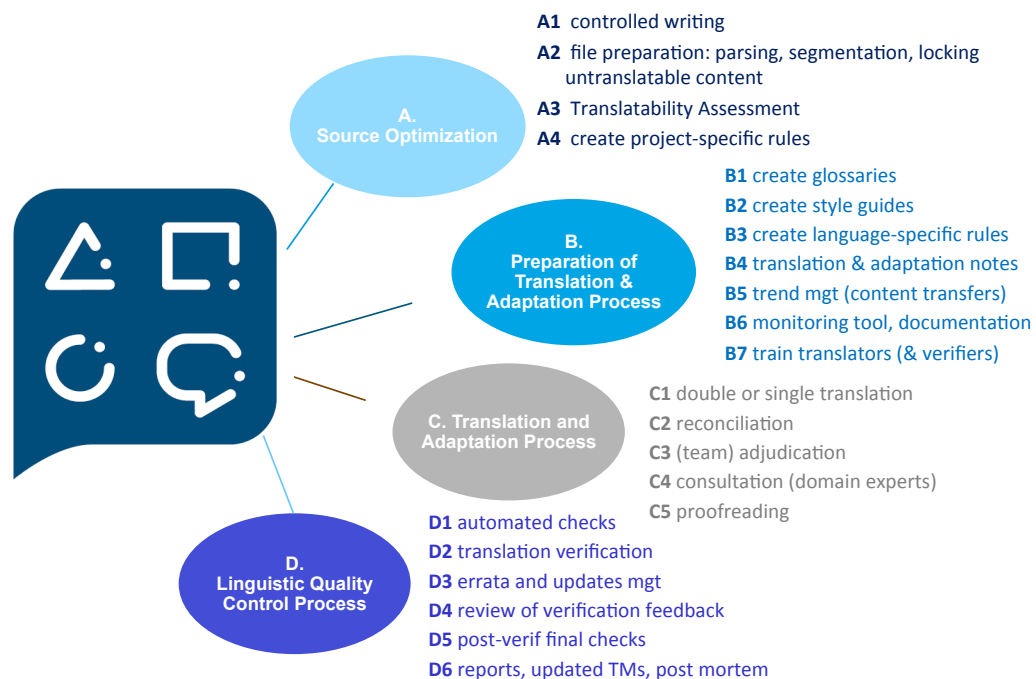
Tension between budget and quality requirements

Insufficient attention given to existing resources on known issues
(new teams want to implement new revisions)

Without thorough project management, multiple revision layers add more time and cost than value

Automated checks save time and increase accuracy – but project-specific rules, glossaries, style guides etc. need to be prepared in advance

Downstream fixes: (much) harder to implement in CBA environment





TRANSLATABILITY ASSESSMENT: METHODOLOGY

Stimuli and items are transferred to an Excel form and submitted to a panel of linguists

Covering a variety of target language groups

Draft translations to identify adaptation hurdles

- Use translatability categories to describe these
- Propose adaptation note to circumvent the problem identified
- Or suggest alternative wording without loss of meaning




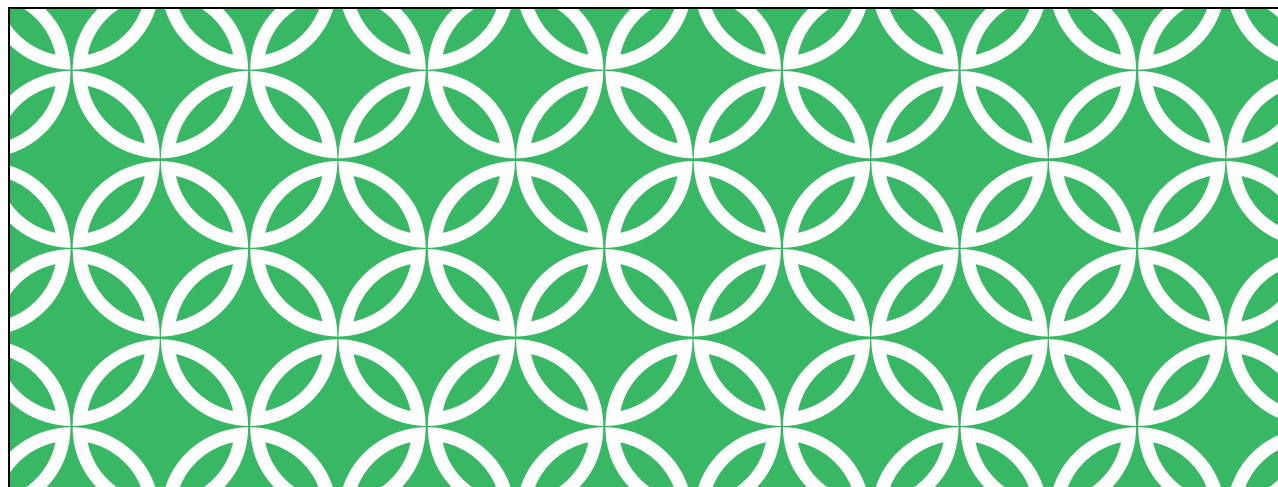
Senior linguist analyzes and collates feedback => consolidated Translatability Report

Location or Section	English	Translatability Evaluation	COMMENT 1
Stimulus Text 1 (4)	Fog harvesting uses mesh of various types of material to collect water from fog.	KNOWN DIFFICULTY, KNOWN WORKAROUNDS	"mesh nets" will be easier to render in all languages and more accurate.
Stimulus Text 1 (5)	Water droplets from the fog are caught on the mesh and collect into larger droplets which flow down into a trough.	STRAIGHTFORWARD KNOWN DIFFICULTY, KNOWN WORKAROUNDS POTENTIAL ISSUES POTENTIALLY AMBIGUOUS UNNECESSARILY COMPLEX REQUIRES REVIEW POTENTIAL CULTURAL ISSUE AGREEMENT ISSUE	collect into": although correct in ENG, the intransitive use of the verb to collect may be tricky and potentially misunderstood (this may be the case for some romance languages), especially when followed by the same verb with transitive use in the next sentence.

COMMENT 1	SUGGESTIONS FOR ALTERNATIVE WORDING
will be easier to render in all languages and more accurate.	Fog harvesting uses mesh nets (...)
although correct in ENG, the use of the verb to collect and potentially misunderstood (this may be the case for some romance languages), especially when followed by the same verb with transitive use in the next sentence.	(...) and gather into(...)



STRAIGHTFORWARD	No potential translation or adaptation problems identified during the advance translation of this segment into languages from at least two language groups.
KNOWN DIFFICULTY, KNOWN WORKAROUNDS	A translation/adaptation difficulty has been recognised in this segment and has been encountered in the past. Satisfactory solutions to this issue have been successfully implemented.
POTENTIAL ISSUES	The current wording or content of this segment is likely to give rise to translation or adaptation problems in some languages, to the extent that functional equivalence may be difficult to achieve.
POTENTIALLY AMBIGUOUS	The current wording or content of this segment could be interpreted in more than one way and it is desirable to disambiguate the source version of this segment before submitting it for translation/adaptation.
UNNECESSARILY COMPLEX	The current wording or syntax of this segment is somewhat contorted, for example due to use of several clauses, questions embedded in questions or unnecessary use of passive voice. The source version can be simplified without loss of meaning.
REQUIRES REVIEW	The current source version of this segment is not suitable for translation/adaptation and needs to be edited before submitting for translation/adaptation.
POTENTIAL CULTURAL ISSUE	The semantic content of this segment may be difficult to adapt in a particular cultural or language group.
DOUBLE-BARRELLED	A question touches upon more than one issue, yet allows only for one answer. Many double-barrelled questions can be detected by the existence of the grammatical conjunction "and" in them.
AGREEMENT ISSUE	There is either an agreement issue within the segment (e.g. subject-verb agreement, or sequence of tenses, or a pronoun-antecedent agreement) or an agreement issue between two segments (e.g. no grammatical match between a question and response options).
CONSISTENCY	In this segment, a different term, expression or form of address has been used versus other occurrences of similar content; and this inconsistency seems to be unintentional.
REDUNDANCY	This segment contains a tautology or unnecessary repetition. Removing it would not alter the meaning of the segment.
POSSIBLE ADDITION	The current wording or syntax of this segment is elliptic or unclear, and its implicit meaning is likely to get lost in translation. This could be solved by adding a word or a piece of information.
LOGICAL PROBLEM	This segment contains a logical problem or there is a logical problem between this segment and another segment, and this issue seems to be unintentional.

CONCLUSIONS

WHAT NEXT?



OUTLOOK

More automated QA for segments not linked to measurement

(combined with human verification of representative sample?)

Focused verification of pre-defined key segments
[contingent on more upstream preparation work]

More work to be done on relation between important segments



OUTLOOK

Translatability of source materials: consider alternatives to the “second source” concept, e.g. multiple “advance translations” feeding back to the source. Cost/timeline/tradeoff factors

Combine DIF analysis with statistics on verification findings

Shift from trial & error to predictive analysis



OUTLOOK

Further develop support to non-professional users of (open source) translation technology so that all players can use the same platforms

Maintain repositories of searchable translation memories (with user management)

Project: cross-verification: e.g. TUR<>KOR and CHN<>RUS



THANK YOU VERY MUCH

steve.dept@capstan.be