

A close-up, high-contrast, black and white photograph of a lion's face, focusing on its eye and the texture of its fur. The image is dark and moody, with the lion's eye being a prominent light feature.

Best practices for assessing test translations during tendering procedures

Attila Görög

May 2019, Brussels



CONTENTS

- Test translations
- Evaluation (MQM-DQF)



Test Translations

Best approach to test translations: NO TEST TRANSLATIONS

- Test translations don't allow LSPs to follow standard translation processes and often force them to break industry best practices.
- Taking an alternative approach is in our best interest: engage in a pilot with selected companies.
- It will save time, minimize risk, and ultimately result in a better process to select the right partner

How do test translations break all the rules?

- There are some best practices that all LSPs follow to give customers the highest quality translations within the context of a real project.
- Most sample translation quality tests do not allow language vendors to adhere to those best practices.



Terminology

- A glossary is a critical component of a quality translation (especially for highly-complex domains such as *gaming, medical, and legal*).
- Best practice is to obtain a current glossary or develop one and submit it for the customer's review and approval before beginning the actual translation.
- **In most sample translation tests, a glossary is not provided, nor is there time to create one!**

Style guides

- They govern tone, usage of terminology, punctuation, sentence structure, date/time/number formats, etc.
- Even if the client only has an English style guide, that provides a place to start.
- The LSP can leverage this to help the client develop style guides for each additional language.
- **A style guide is rarely, if ever, provided for test translations!**

Training

- Very important to translators
- The companies that receive the highest quality translations from their LSPs invest in training the vendor's resources
- **Sample translation tests rarely ever have associated training!**

Q&A

- During a normal translation process, translators might ask questions about source content to make sure they understand it correctly.
- Customers clarify terms, meaning, intent, etc., and translators revise or construct the translation accordingly.
- **During the sample translation test process, Q&A is rare; therefore, translators are left to guess!**

Subject Matter Experts (SMEs)

- Most test translations have specialized content—meaning that it’s the most difficult to translate in a test setting.
- To provide the highest quality translations, LSPs work to find translators who are SMEs in the given area.
- Sample translation tests assume the best resources are available “on-demand” but these expert resources are often booked for other projects.
- **SMEs are removed from paid projects to address a sample that requires a quick turnaround, breaking all best practice rules.**

Speed

- Often, clients request that test translations be completed in 2 or 3 days. This simply does not allow for all the optimal steps to take place.
- In the context of an established program, the vendor would have client-familiar resources already in place and would be able to act quickly.
- It is nearly impossible to engage the right resources when adequate lead time is not given, and when translations are due back immediately.
- **Kickoffs and other information exchanges are skipped, as well as the Q&As between the translators and the client.**

To sum up



Terminology

Style guides

Q&A

Training

SMEs

Speed

Purpose

- Why is the test translation needed in the first place?
- Is it to test the ability of an LSP to **act quickly**?
- Is it to test their ability to **communicate** with you?
- Is it to see how the LSP **operates**?
- Is it to assess the **quality** of the final product?



None of these goals can be judged effectively via a test translation.

The customer has asked for test translation – *WHAT NOW?!*

- Find out what the customer means by “test translation”
 - *Trial or pilot* – usually paid, open dialogue with customer
 - *Test translation* – normally not paid, competitive
- Aim for trial or pilot project instead of test translation
- If customer insists on a test translation, negotiate scope
 - *Number of languages*
 - *Size of content*



Some questions to discuss

- Have you been involved in a tender involving test translations?
- If yes, what was your experience?
- Why are test translations needed?
- Are there alternative ways to assess an LSP?
- How can we optimize test translations?



Alternative testing method



Samples

References

Quality

Resources

Project Management

Samples

- Ask for **samples of work** from similar customers to assess the translation supplier's quality in a particular domain.
- By providing materials already translated in the customer's domain, that comply with the LSP's best practices, the customer will be able to more realistically review and evaluate the LSP's work.

References

- Ask the LSP for **customer references**.
- A good LSP will be able to showcase quality and expertise by providing testimonials and use cases of previous projects.

Quality

- Ask for in-depth information about **quality steps**
- Does the LSP have documented processes for each of these? Do they follow their own process consistently?

Resources

- Be clear about your perceived **success criteria**.
- Ensure the LSP understands what resources you feel are required to succeed on your project, and ask for information on resourcing that validates the LSP is able to secure the right resources.

Project Management

- Ask about **project management**. How does the LSP's PM control quality?
- If you believe it is critical to assess an LSP's work on a live job, conduct a small, paid, **pilot project**.

To sum up

- You can control the time and cost of evaluating the materials by short listing language providers and **asking only two** to complete a pilot
- Ask these LSPs to complete translations that you will actually use, so your money is not spent on **redoing and evaluating** already translated material.

Evaluation (MQM-DQF)



What is a quality Translation?

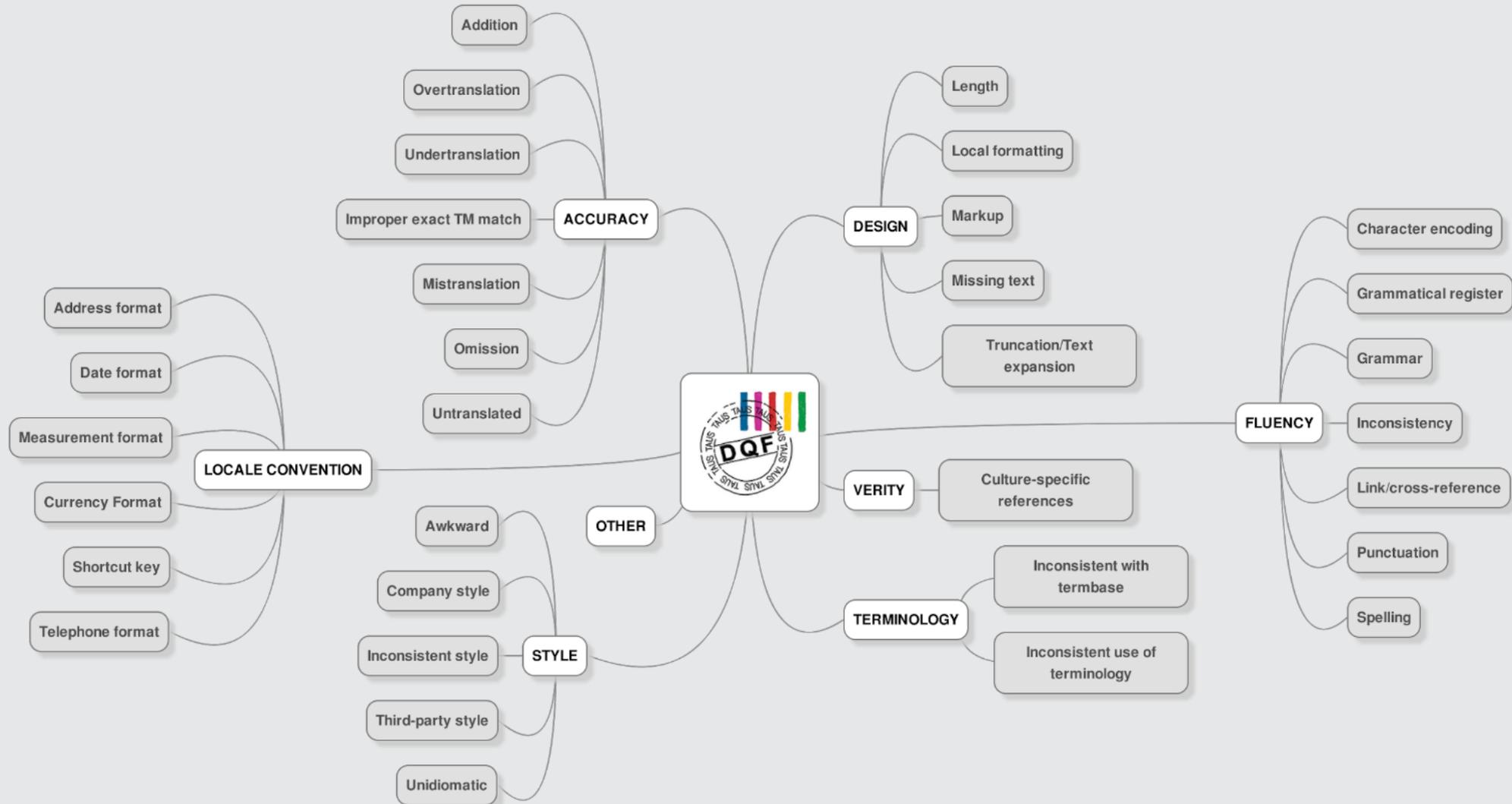
- A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account requester and end-user needs (ASTM F2575).

What do we mean by specifications

- Cover all aspects of translation projects
 - Quality
 - Process
 - Project Management
 - Turnaround time
- Focus today is on the “final product”
 - Source-content information
 - Target-text requirements



DQF-MQM Error Typology



What is DQF-MQM

- A **taxonomy** of translation errors
- A **standard** for benchmarking quality
- A **process** of creating **analytic metrics** that share a common basis with specifications
- A **scoring method** to quantify translation quality
- A **collection of metrics** relevant to any sort of translation as well as evaluation of source-text quality

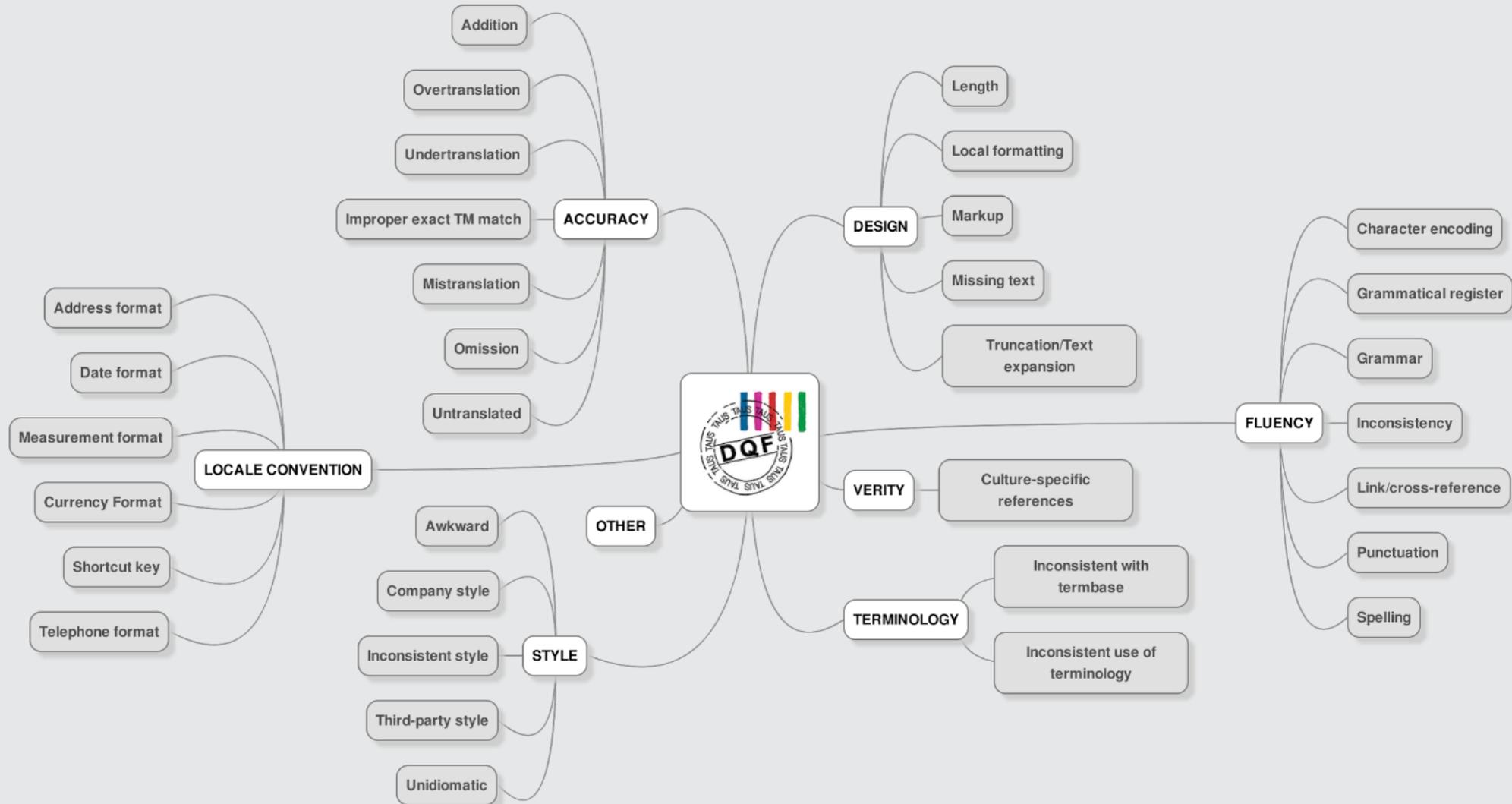
History

- Translation quality assessment was subjective
- 1990s: move to vendor-specific checklist-based criteria for quality assessment based on error counts
 - LISA QA Model
 - SAE J2450 (automotive industry)
- Often heavily modified
- Attempt in ISO to implement a universal metric for all translation (cancelled in 2013)

Harmonization

- There was a need for a standard method that offered flexibility
- DQF and MQM error typologies were harmonized under QT21 in 2015
- Three person team from TAUS, DFKI & LTAC
- DQF-MQM is the short name for the subset of MQM error typology that is harmonized with TAUS DQF

DQF-MQM Error Typology





Accuracy

- An accuracy error means that the translation does not reflect the source.
- The target text refers to matters that are not referred to in the source, or added, neglected or changed some of the meaning of the source.
- Example: Translating the Italian word 'canali' into English as 'canals' instead of 'channels'.

Locale Convention

- An error that neglects conventions or rules in the target language.
- This can be the way the dates or addresses are written in the target language, certain styles of punctuation.
- Example: A text translated into JP uses Western quotation marks rather than the appropriate JP quotation marks (「 and 」).



Verity (Locale Appropriateness)

- A verity error makes a false statement about the outside world.
- Example: it doesn't correctly list the opening hours of a shop.





Fluency (Linguistic Well-formedness)

- As a monolingual error type, fluency takes into account whether the target language text is formed correctly.
- Spelling and grammar errors are the most typical ones in this category.



Style

- Errors caused by the style of the target text.
- Even though the text can be fluent, the style can be awkward, inconsistent or not according to a style guide.
- Example: The translation of a light-hearted and humorous advertising campaign is in a serious and “heavy” style even though specifications said it should match the style of the source text.



Design

- Errors associated with layout issues.
- Problems with the formatting of the text causes issues in this type of error.
- Example: an incorrectly formatted document



Terminology

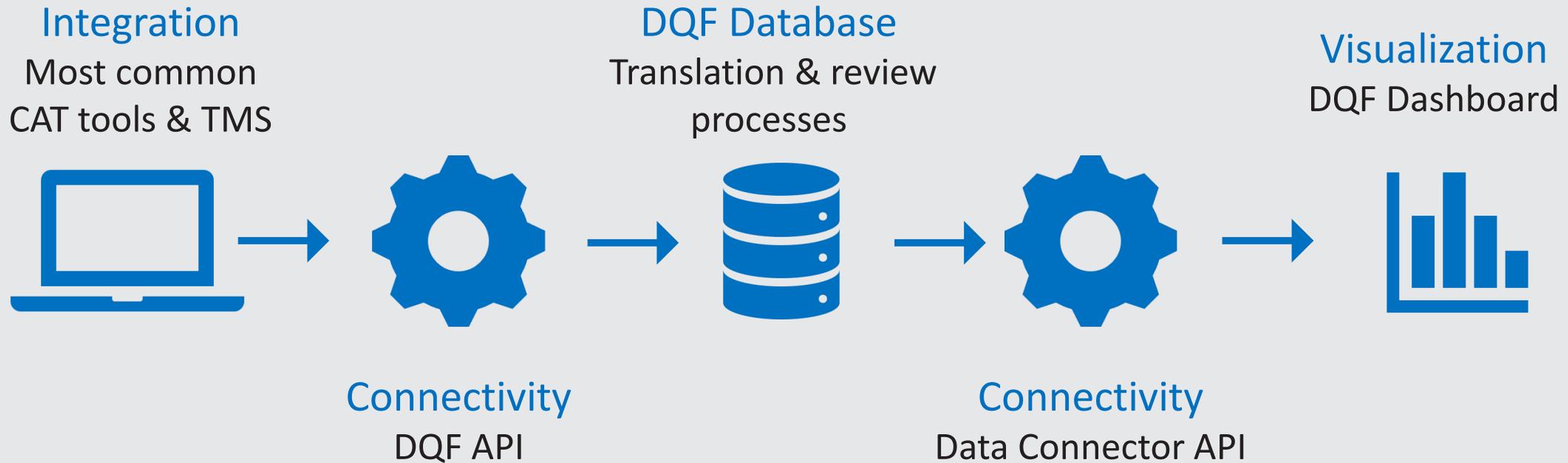
- Errors caused by the wrong choice of a term in the target language.
- This can be caused by neglecting a terminology database, or by using terminology in an inconsistent way.
- Example: the English word *e-mail* is translated as *e-mail* in French, while the guidelines mandated that *courriel* be used.



Other

- This category serves as a safety net in case the error can not be described by any of the other categories.
- Depending on the error annotation environment, some extra explanation can be added regarding the annotated error, but the ‘Other’ category, for all statistical purposes, is treated just like any of the other error categories.

How TAUS DQF Works



Step 1: define the review settings

Review Settings

Select the Review Settings you want to apply to your DQF project. This can also be done later.



Select Review Type (Selecting 'None' will not apply any review settings)

None Correction Error Annotation Combined (Correction and Error Annotation)

Error Categories

- Accuracy
 - Addition
 - Omission
 - Mistranslation
 - Over-translation
 - Under-translation
 - Untranslated
 - Improper exact TM match
- Fluency
 - Punctuation
 - Spelling
 - Grammar
 - Grammatical register
 - Inconsistency
 - Link/cross-reference
 - Character encoding
- Terminology
 - Inconsistent with tembase

Additional Properties

Error Severities

	Name	Penalty Points
▶	neutral	0
	minor	1
	major	5
	critical	20

Pass/Fail Threshold (per 1000 words)

20

Sampling%

0



Step 2: apply the error typology during the review

ABC source text.txt.sdlxliff [Review]*

1	This is sentence number 1.		Šis ir pirmais teikums.
2	This is sentence number 2.		Šis ir otrais teikums.
3	This is sentence number 3.		Šis ir trešais teikums.

ABC source text.txt

Error Categories

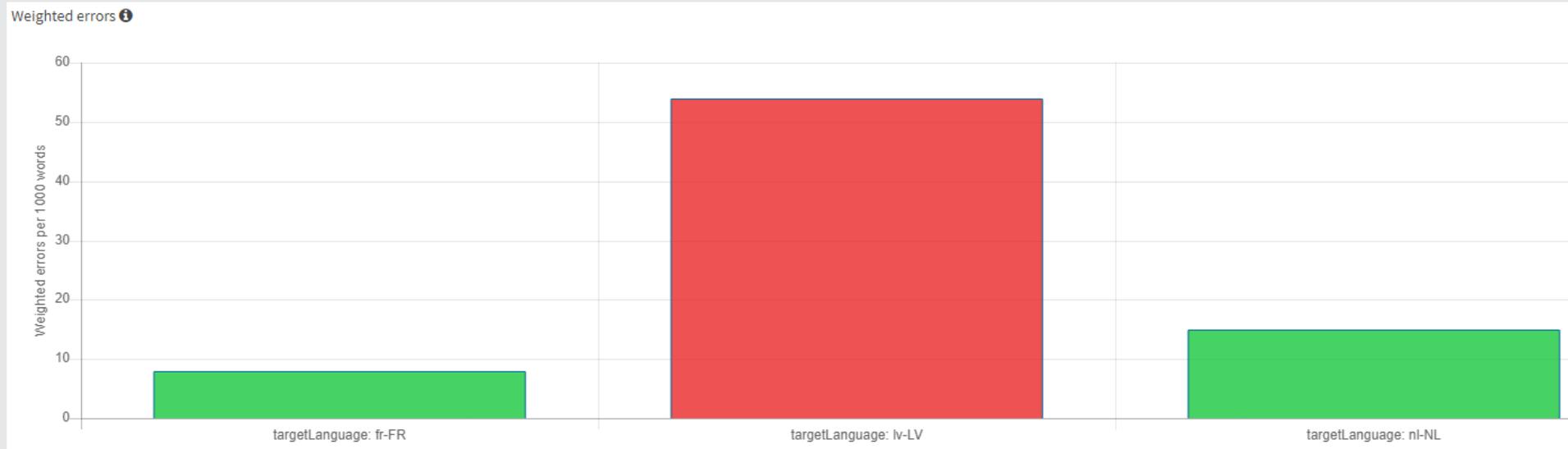
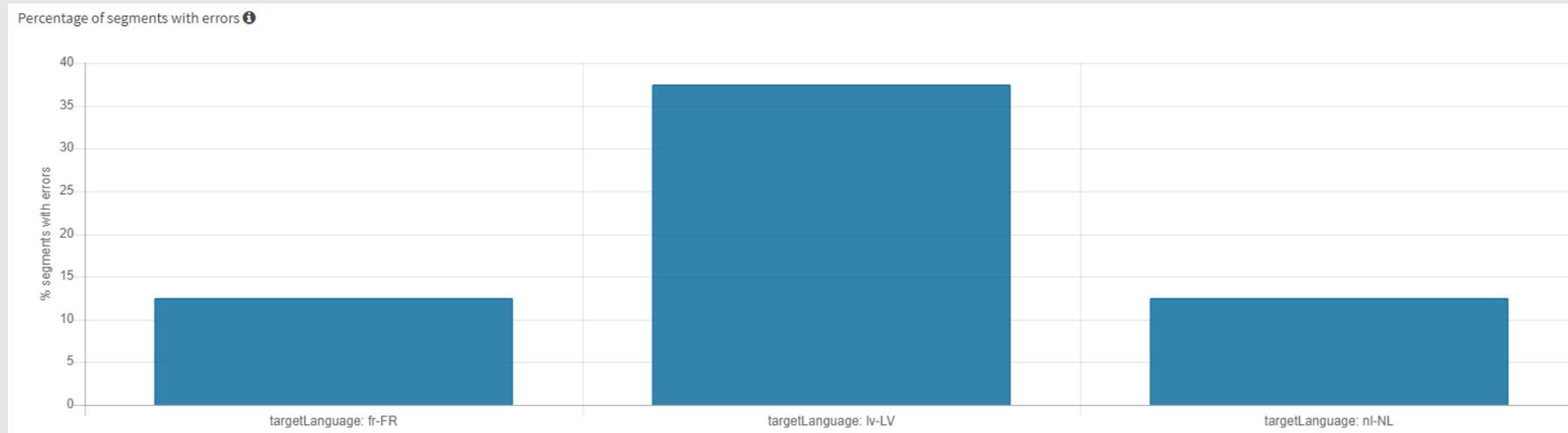
		Error Category	Shortcut	Severity	Error Category
		Accuracy	Ctrl + 1	minor	
		Addition	Ctrl + 11	neutral	
▶		Omission	Ctrl + 12	neutral	
		Mistranslation	Ctrl + 13	neutral	
		Over-translation	Ctrl + 14	major	
		Under-translation	Ctrl + 15	critical	
		Untranslated	Ctrl + 16	neutral	
		Fluency	Ctrl + 2	neutral	
		Punctuation	Ctrl + 21	neutral	
		Spelling	Ctrl + 22	neutral	

Step 3: track the review results in the CAT tool

The screenshot displays a CAT tool interface. The top window, titled "ABC source text.txt.sdlxliff [Review]", shows a list of three sentences for review. The third sentence is selected. The bottom window, titled "Overview", shows a table of review results.

	Segment ID	Date	Document	Author	Severity	Error Category	Applied To
▶	1	18-Jan-19 15:50	ABC source text.tx...	Revie	minor	Accuracy	pimais
	3	18-Jan-19 15:55	ABC source text.tx...	Revie	major	Over-translation	*** Whole Segment ***
	3	18-Jan-19 15:55	ABC source text.tx...	Revie	minor	Spelling	Šis

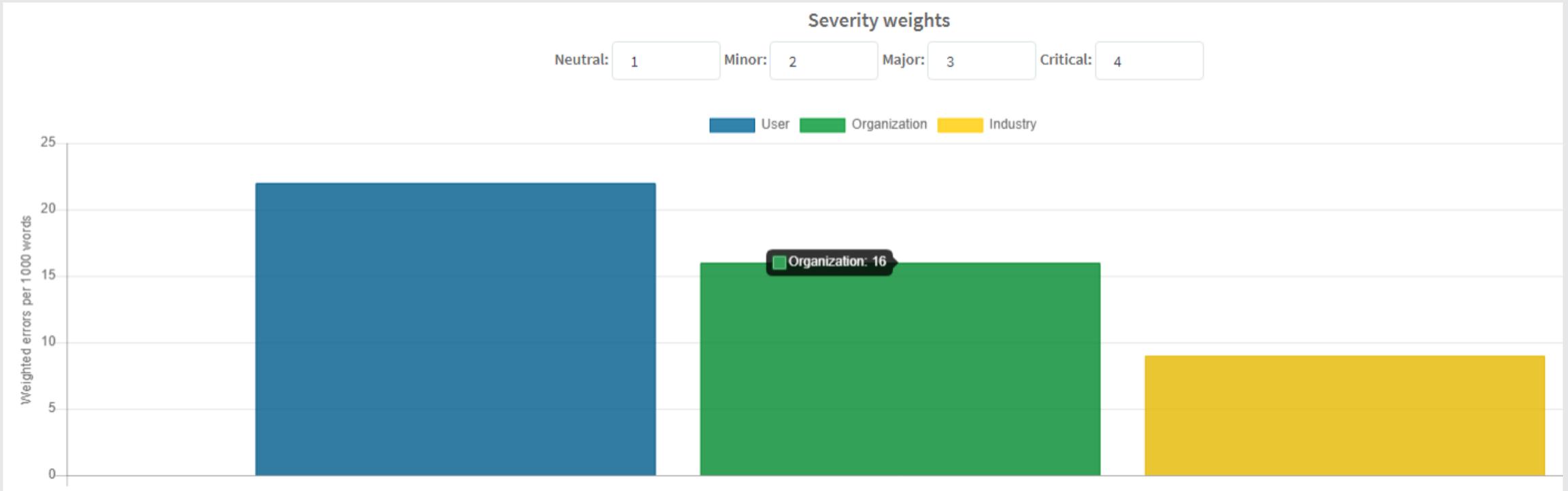
Step 4: analyze the data on the DQF Dashboard



Step 5: download segment-level report

Target lang	Segment orig	Source segment	Translated target segment	Subsegment	Error position start	Error position end	Error category	Error severity	Penalty points
nl-NL	TM	Modern smartphones have a touchscreen color display with a graphical user interface that covers the front surface and enables the user to use a virtual keyboard to type and press onscreen icons.	Moderne smartphones hebben een Kleurendisplay met tipscherm met een grafische gebruikersinterface die bedekt de voorste oppervlak en kan de gebruiker een virtuele toetsenbord en drukt u op het scherm pictogrammen.	Kleurendisplay	31	45	Style	minor	2
fr-FR	TM	Smartphones function using a rechargeable lithium-ion battery.	En utilisant la fonction de smartphones une batterie lithium-ion.	En utilisant la fonction de smartphones une batterie lithium-ion.	-	-	Fluency	neutral	1
lv-LV	TM	It possesses extensive computing capabilities, including access to the Internet using both Wi-Fi and mobile broadband.	Tam piemīt plaša skaitļošanas tehnikas iespējas, tostarp interneta pieslēgums, izmantojot bezvadu un mobilo slū.	plaša skaitļošanas tehnikas	11	38	Accuracy	major	3
lv-LV	TM	Modern smartphones have a touchscreen color display with a graphical user interface that covers the front surface and enables the user to use a virtual keyboard to type and press onscreen icons.	Mūsdienu viedtālruni have a touchscreen krāsu displejs ar grafisku lietotāja saskarni, kas aptver priekšējā virsma un ļauj lietotājam izmantot virtuālā tastatūru, lai tipa un preses lezīmēšana ekrānā ikonām.	a touchscreen krāsu displejs	26	53	Terminology	critical	4
lv-LV	MT	A high-capacity portable battery charger.	Lieljaudas pārnēsājams akumulatora lādētājs.	Lieljaudas pārnēsājams akumulatora lādētājs.	-	-	Locale convention	neutral	1

Step 6: benchmark against the industry



Key Takeaways

- Translation quality is well-defined but not static
- Translation specifications matter a lot
- DQF-MQM provides a standardized yet flexible way to evaluate translation quality
- Translation quality can be quantified and benchmarked
- Collect data and start benchmarking (through TAUS API or scorecard)
- It's essential to train and assess evaluators

Discussion

- How do you evaluate translations right now?
- Are there alternative methods in use in the industry?
- How can we as a group make sure test translations and their evaluations follow certain guidelines?



Thank you.

For more information please contact:

Attila Görög

Account Executive Lionbridge

+31 6 34 103 280

Attila.gorog@Lionbridge.com

www.lionbrige.com