

Measuring Translation Quality with Smartling

We aren't just confident in the quality of our translations. We guarantee it.



© 2023 by Smartling, Inc.

Smartling is committed to delivering best-in-class results to businesses with a multilingual footprint. We are not just confident in the quality of our translations. We guarantee it.

Learn more about Smartling's approach to quality, how we measure it, and how we ensure that our customers consistently receive high-quality translations across all workflows and languages — with zero risk.

Measuring translation quality, much less guaranteeing satisfaction with high-quality work, has been a longstanding challenge for those in localization.

While a translation can have objective issues — typos, grammar mistakes, or even blatant mistranslations, omissions, or additions — language is also personal. Ten expert translators can take the same source text and produce ten different yet “excellent” translations, each informed by their unique use of the target language. Those linguistic preferences can make evaluating translation quality a less-than-straightforward endeavor.

Absent proper safeguards, translation assessments often slip into the subjective territory of “like/dislike.” Trusting the results of a single evaluation, let alone monitoring translation quality over time across multiple content types and languages, becomes impossible. Ultimately, that has profound implications for businesses looking to scale their multilingual footprint. After all, translation quality issues can significantly impact a company’s brand image, reputation, and

Intro

bottom line. It is only reasonable that businesses expect their chosen language services provider (LSP) to deliver on their promise of high quality, both the first time and every time after that.

At Smartling, we pride ourselves on consistently delivering world-class translations — and doing so while keeping time and cost constraints top of mind. But we don't just expect you to take our word for it.

We can prove it.

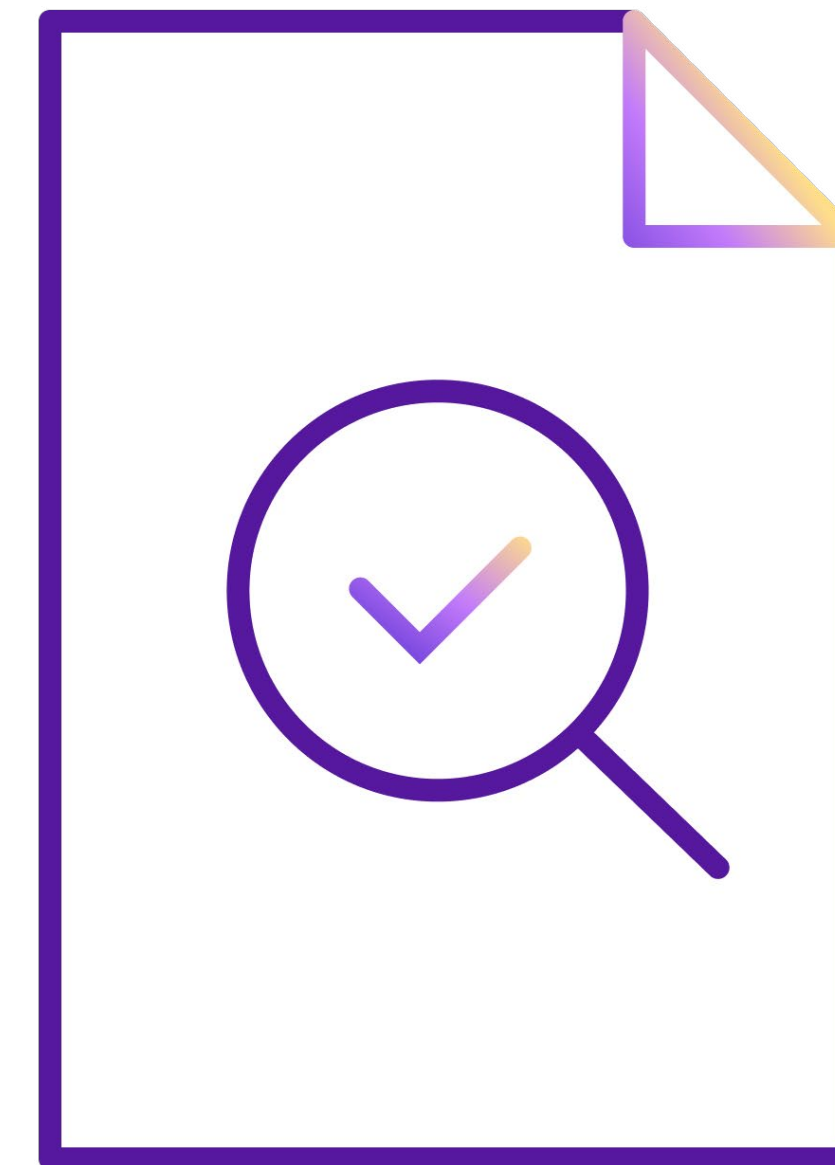
This report explores our repeatable approach to measuring translation quality. It then explains how Smartling ensures our customers consistently receive high-quality translations across all workflows.

Finally, we offer some additional food for thought to help you jumpstart a deeper conversation about translation quality with your current LSP — whether that's Smartling or another provider.

How we measure translation quality

We designed our quality evaluation program around a simple question:

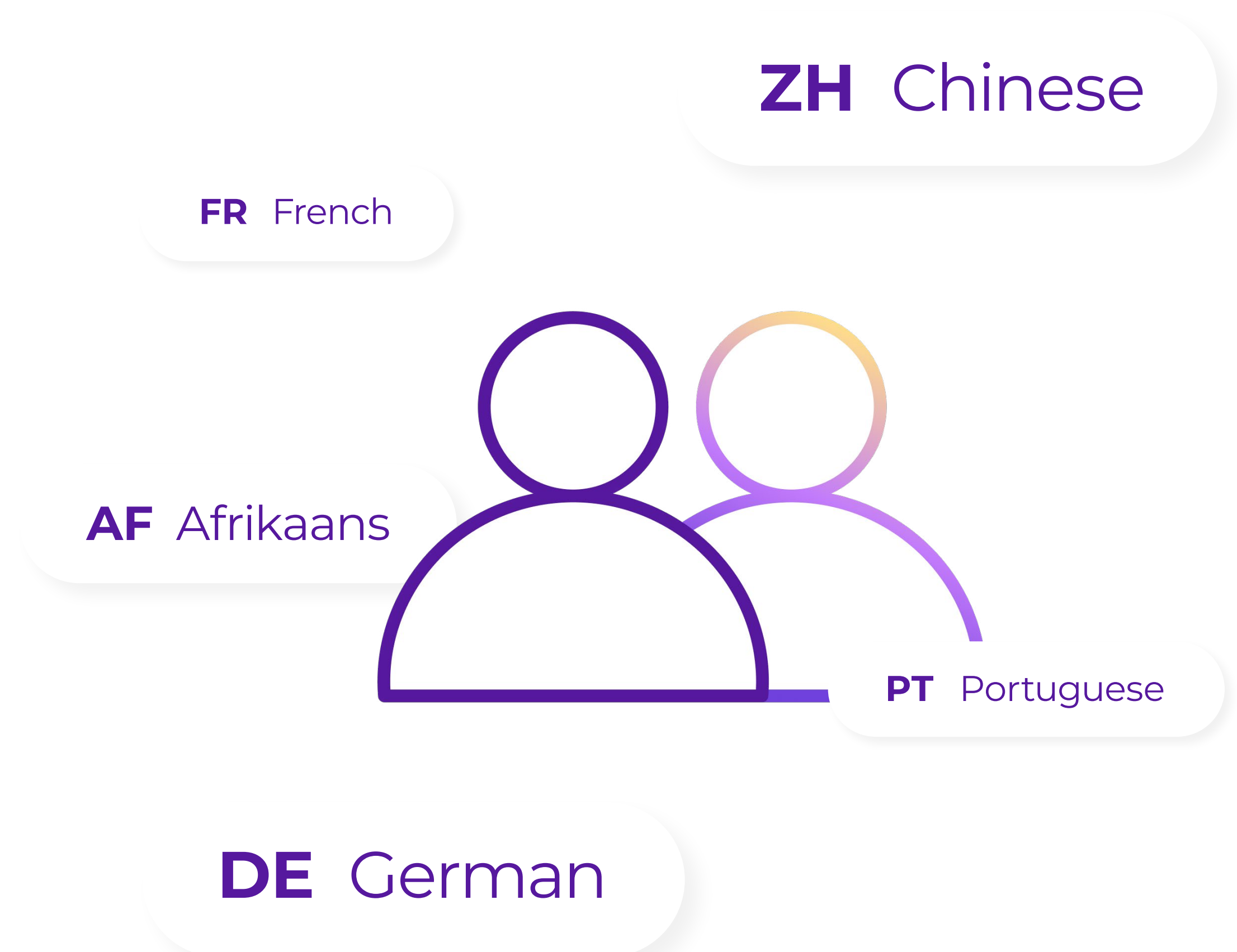
Can we guarantee high-quality translations using our LanguageAI™ platform?



Smartling customers are as varied as can be

They operate across many different industries and use our LanguageAI platform to translate all types of content. Some may translate their content into one or two other languages, while others translate into upwards of twenty or thirty languages.

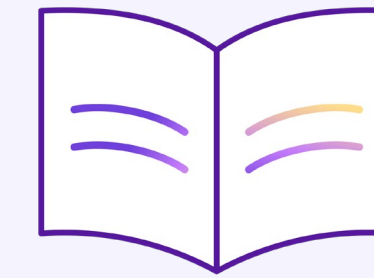
In addition, some customers prefer that their content go through a human-only translation workflow. Others opt for machine translation (MT) or Smartling's AI-powered human translation workflow that brings together the best of humans and technology. Still others choose to do a mix, relying on Smartling's LanguageAI platform to route their content to different translation solutions depending on the content type and their business needs. Therefore, our evaluation process needed to set us up to address all of these scenarios.



What is MQM?

Language is subjective. Individual preferences can influence whether someone views a translation as “good” or “bad.” That, in turn, makes it challenging to measure translation quality consistently. To safeguard against this, having a clear definition of quality is a must. That includes establishing uniform criteria that evaluators should apply to all translations.

MQM provides a framework for measuring translation quality that focuses on identifying and quantifying errors in a text. Each error is recorded, categorized, and assigned a severity level — always with the individual customer’s translation specifications (e.g., intended purpose, terminology requirements, and style preferences) in mind.



**Multidimensional
Quality Metrics (MQM)**
is a framework for measuring
translation quality that focuses
on identifying and quantifying
errors in a text.

Error categories

Accuracy

Fluency

Style

Technical/Local

Errors fall within overarching categories, such as Accuracy, Fluency, Style, and Technical/Locale. Each category contains a list of error types. For example, under the umbrella category of Accuracy, an evaluator may find the following error types: Addition, Omission, and Mistranslation. It is important to note that MQM has several overarching categories and many more error types. It is up to individual users to select only those categories and types most suitable for their purposes when designing a quality evaluation program.

Severity levels (e.g., critical, major, minor, neutral, etc.) help determine the impact of each error on the overall translation. Evaluators must rate the extent to which the error affects the intended purpose of the text.

Once all the strings have been reviewed, the errors are weighted according to the framework. Afterward, a score is calculated by normalizing the data set against using a standard volume (word count).

MQM scorecard: Top-level error typology with 4 severity levels

	Error counts				Error type penalty total	
Severity penalty levels:	Neutral	Minor	Major	Critical	ET Weights	ETPTS
Severity penalty multipliers:	0	1	5	25		
Error types						
Terminology	2	3	1	0	1.0	8.0
Accuracy	4	5	2	0	1.0	15.0
Linguistic conventions	1	8	0	0	1.0	8.0
Style	5	3	0	0	1.0	3.0
Locale convention	1	2	0	0	1.0	2.0
Audience appropriateness	0	2	0	0	1.0	2.0
Design and markup	0	6	0	0	1.0	6.0
Custom	0	0	0	0	1.0	0.0
					Absolute penalty total (APT): 44.0	
					Evaluation word count: 10184	
					Penalty per 100 words: 0.43	
					Overall score: 99.57	
					Decision: PASS	

Why MQM

MQM was developed in 2012–2014 as part of the European Union-funded research project QTLaunchPad. It was refined in 2015–2018 with the QT21 project, spearheaded by the German Research Center for Artificial Intelligence (DFKI). While it may seem to be a newcomer to the translation quality assessment landscape, MQM draws upon, harmonizes, and improves upon several older QA methodologies, including the LISA QA model, SAE J2450, and the TAUS Dynamic Quality Framework.

The beauty of MQM lies in its flexibility and repeatability. Previous methodologies were introduced with specific purposes and content types in mind. In contrast, the MQM framework is decidedly not one-size-fits-all.



MQM quick facts:

- Developed in 2012–2014
- Refined in 2015–2018
- Flexible and repeatable methodology

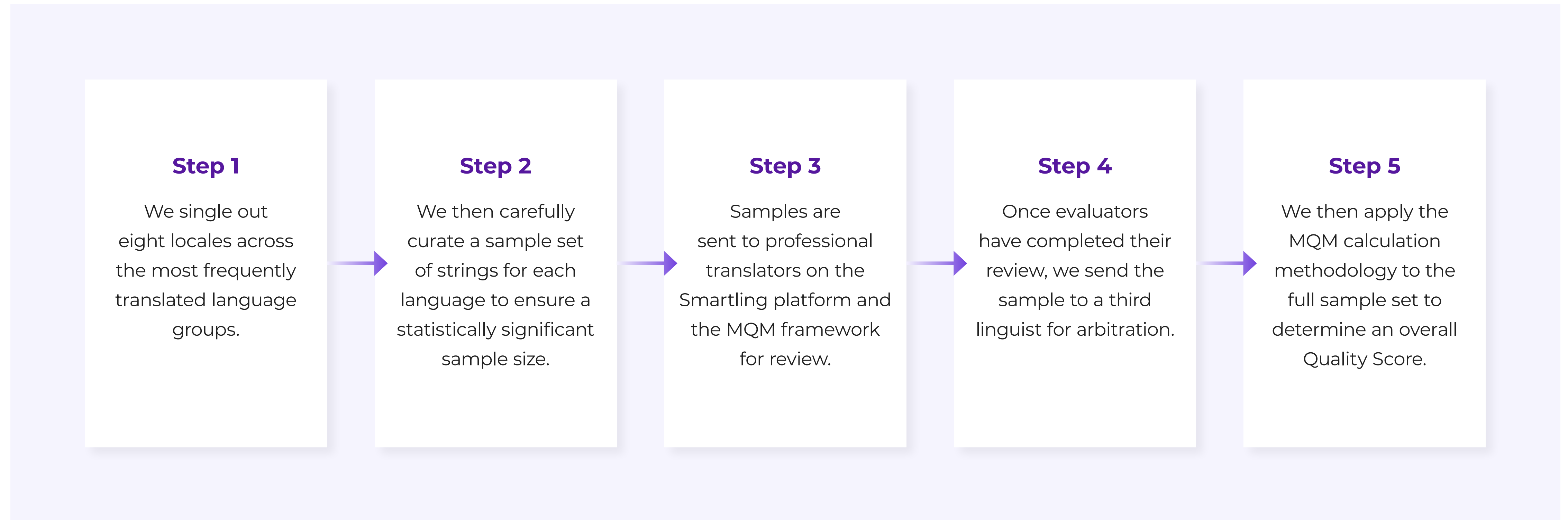
"Implementing MQM at Smartling has enabled us to ensure the high quality of our final translations across all workflows."

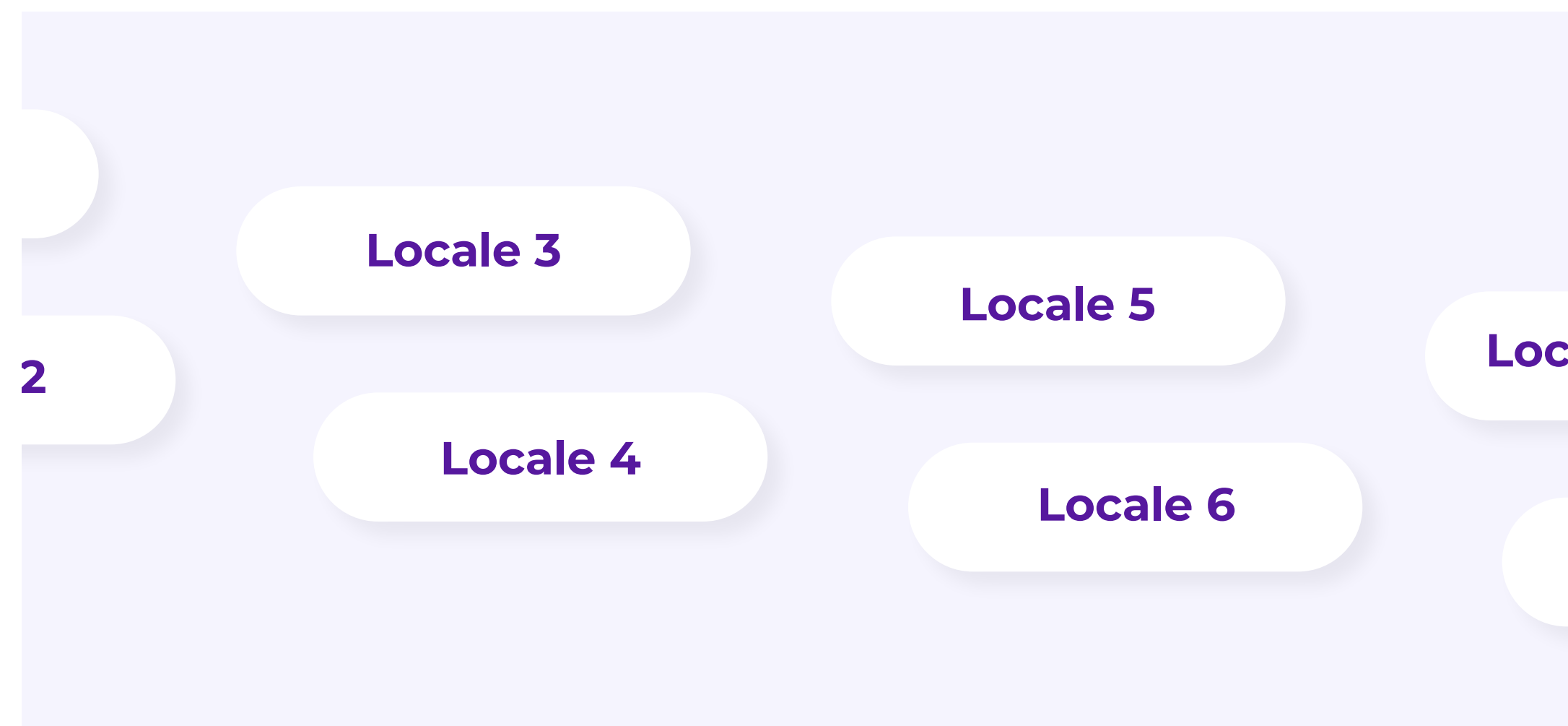
In practice, this means that MQM can be used to evaluate not only a translation, but also the source text. That can help identify potential problem areas before translation.

The MQM framework also bridges the gap between assessing human and machine output. It can apply to human-only translations and those produced with machine, or AI, involvement. That allows for effective comparison.

Implementing MQM at Smartling has enabled us to ensure the high quality of our final translations across all workflows. We can also compare the performance of our translation solutions to more easily discover actionable insights. Ultimately, that helps us to monitor translation quality across our customer base, conduct root cause analysis, and establish a consistent improvement process.

Smartling's quality evaluation process





Step 1

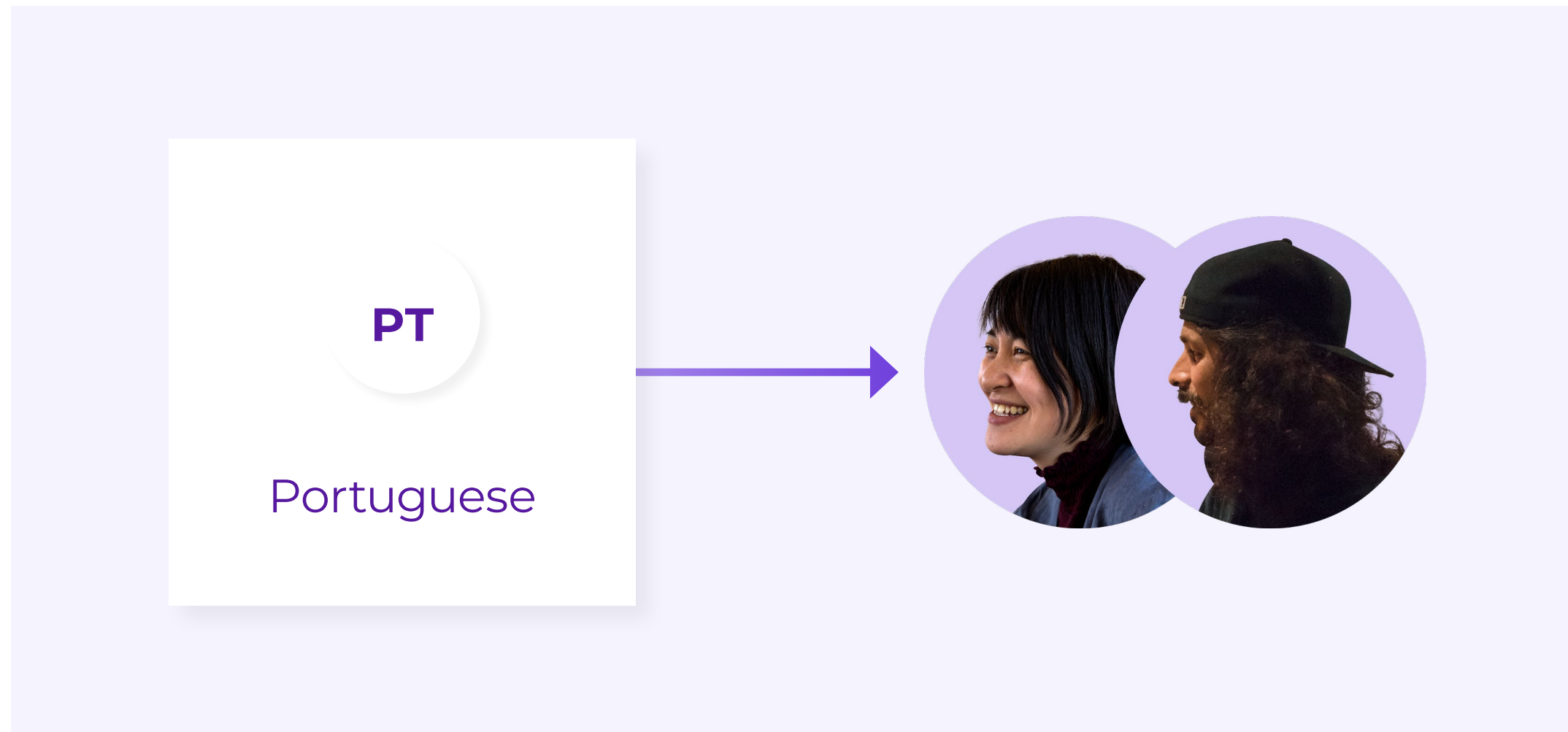
We single out eight locales across the most frequently translated language groups — e.g., Romance, Germanic, Nordic, Asia-Pacific, Slavic, etc. — for assessment.



Step 2

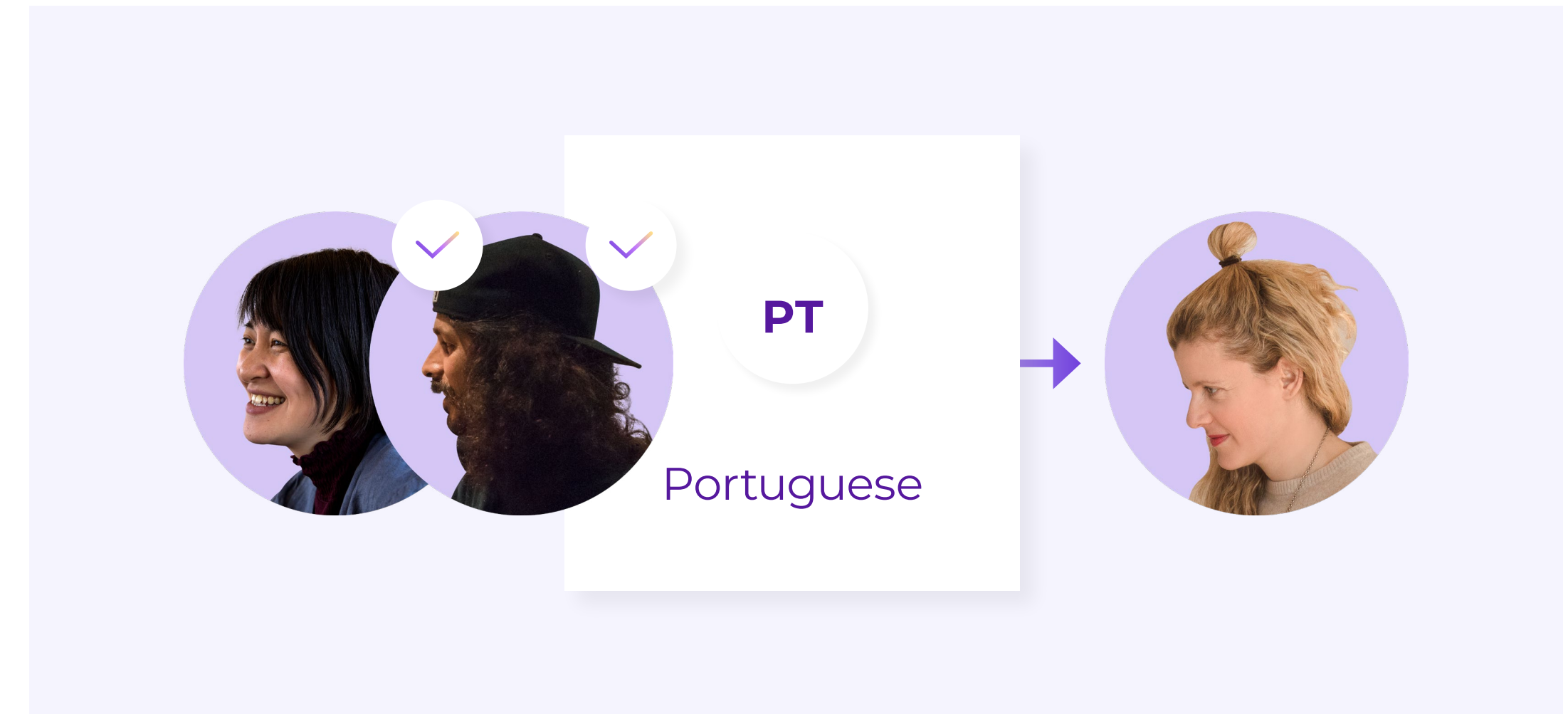
We then carefully curate a sample set of strings for each language to ensure a statistically significant sample size (a minimum of 3,000 words) representative of Smartling customers' content. It makes little sense, for instance, to choose a 100-word sample set of short, grammatically simple strings that don't contain enough of a customer's preferred terminology.

How we measure translation quality



Step 3

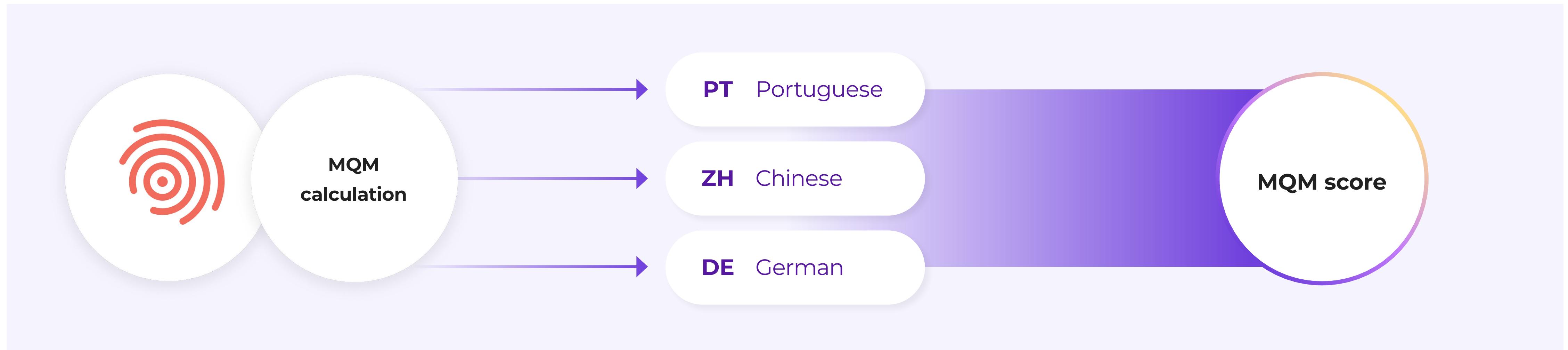
Samples are sent to professional translators with extensive training on the Smartling platform and the MQM framework. To reduce subjectivity, none of our evaluators are part of our regular pool of translators — that is, they have been exclusively recruited for these monthly evaluations — and we engage two evaluators per language.



Step 4

Once evaluators have completed their review, we send the sample to a third linguist for arbitration. This step involves reviewing only the discrepancies between the original evaluators' work. If one evaluator recorded an error, for example, but the other did not, it is up to the arbitrator to determine which evaluator was right.

How we measure translation quality



Step 5

We then apply the MQM calculation methodology to the full sample set to determine an overall Quality Score. We apply the same framework and weights to human-only translation and to automated translation output (e.g., MT).

The result is a comprehensive report detailing MQM scores by language and translation type. This report can be compared to historical results to identify trends and provides insights to guide actions that will improve quality for the future.

Using the data to take action

After the quality assessments are in, we form an action plan based on the evaluators' feedback. This plan is revisited monthly and may entail focusing on specific languages and content, training or sometimes replacing linguists, and revising our guidelines to address specific error types.

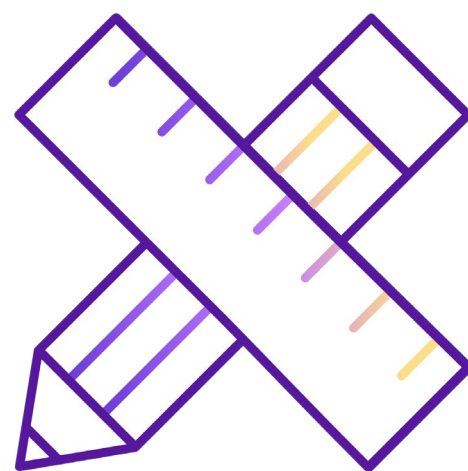
In addition, the MQM scores derived from the evaluations are valuable for training Smartling's machine-learning applications. The human-labeled data serves as a gold standard to test the accuracy of our automated MQM predictions. The data also helps fine-tune large language models (LLMs) for quality estimation and evaluation tasks, as it provides examples of strings (or sentences) annotated by human evaluators for different error types.

Arbitrated evaluations play a pivotal role in this. The trained model does not rely on just one person's opinion on whether an error is minor, major, or critical. Nor does it receive an average score calculated across several evaluators. Instead, it receives one definitive assessment, which teaches the model to differentiate between the good and the bad.

Finally, since Smartling hosts and processes translations for businesses large and small across many different industries, we are in a position to capture several valuable metrics, including: Levenshtein Edit Distance, Translation Error Rate, and Translation Velocity.

How we measure translation quality

We use these, the monthly MQM scores, and other NLP quality estimation metrics to get a well-rounded picture of translation quality. That helps us build and regularly recalibrate our proprietary quality assessment tools.



Levenshtein Edit Distance

Measures the number of changes made to a string by comparing a previous version to the edited version.



Translation Error Rate

Predicts the number of edits needed by comparing the current version of a translated string (e.g., raw MT output) with a reference translation.



Translation Velocity

Measures the time a string takes to move through the workflow, from authorization for translation to completion.

Smartling’s MQM scores

A review of our average MQM scores is revealing. Smartling consistently achieves high MQM scores across all translation workflows. It is important to note that these scores take into account all locales we evaluate. So, Smartling not only achieves high MQM scores for one or two languages but does so across the board.

To put these scores into context, the industry average MQM scores for human translation and traditional MTPE provided by LSPs range between 95% and 97%. In addition, the the industry average MQM score for MT ranges between 87% and 90%. Smartling’s raw MT output consistently scores on the higher end of that range.



AI-Powered Human Translation (AIHT)

Smartling's AI-Powered Human Translation consistently achieves a 98+ MQM score, in-line with human parity, while reducing translation costs and turnaround time by 50% compared to human translation.



Human Translation (HT)

Smartling's Human Translation consistently achieves 98+ MQM scores. Compare this to the industry average of 95-97% for similar workflows.



Machine Translation (MT)

Smartling's Machine Translation consistently achieves 90+ MQM scores. Compare this to the industry average of 87-90%.

How we measure translation quality

Through our LanguageAI™ technology, Smartling is able to deliver high-quality translations that regularly outperform the industry benchmarks. In particular, our AI-Powered Human Translation consistently achieves a 98+ MQM score, in-line with human parity, while reducing translation costs compared to human translation by 50%.

That is why we can confidently say that Smartling guarantees the quality of our translations.

Moving forward, our objective is to preserve our average MQM score of 98% using LanguageAI. We also aim to improve the scores of our raw MT output so that it reaches human-parity quality.

Through our LanguageAI™ technology, Smartling is able to deliver high-quality translations that regularly outperform the industry benchmarks.

What does a quality guarantee mean?

Smartling strives to deliver excellence at every touchpoint. And we are committed to ensuring our customers are satisfied.

To that end, we've designed a comprehensive quality management system. In addition, Smartling regularly undergoes external third-party auditing to provide another layer of quality assurance and to help us fine-tune our internal processes.

As a result, we are confident that the translations we deliver to our customers are of the highest quality. And that is backed by a guarantee: ***If you're not satisfied with your translation, we will rework it at our cost.***

100%

Translation quality guaranteed



Smartling's quality evaluation toolkit

Smartling customers who take advantage of our expert language services can rest assured that we regularly monitor our translation quality.

Quality evaluation is built into our internal processes. There is no additional cost nor action on our customers' part required.



Quality toolkit includes:

Error Typology

Provides the framework for translation quality evaluation and is based on MQM typology

Quality Services

Allows evaluators to log errors as they review translations

Quality Arbitration

Evaluators and linguists collaborate to produce best translation output

Quality Dashboard

Comprehensive display of results, including quality scores

Smartling's quality evaluation toolkit

But we know there are those who want — or need — to be a little more hands-on with translation quality, particularly if they have internal teams or other LSPs handling specific language pairs or content types. Our quality evaluation toolkit allows you to manage translation quality how it best suits you.

Smartling's in-platform **error typology tool**, which is based on the MQM typology, is fully customizable. To start, customers choose between two templates (Full MQM or Simplified MQM). But they can modify those as needed, adding or removing error categories and types. Customers can also keep the MQM-suggested weights or apply their preferred weights to each severity level.

Those who wish to use our **quality evaluation services** gain access to Smartling's network of quality evaluators. Evaluators review translations and log errors directly in Smartling's in-platform CAT tool. That way, they are working in the same environment as the

translators and have at their fingertips everything they may need: context, linguistic assets (translation memory, glossaries, style guide, etc.), and automated QA checks. Evaluations can be done as part of a production workflow, before the translations are published, or after the fact.

The **arbitration tool** fosters collaboration between the evaluators and the linguists who worked on the content. There is an opportunity to discuss errors and provide justification or feedback, which strengthens feedback loops and improves translation quality moving forward.

Finally, with Smartling, customers can glean actionable insights easily. They can download the raw data themselves. Or, they can view the results — broken down by language, project, or job — via our **Quality Dashboard**, which displays the information (including the final quality scores) in a digestible format in near real time.

Questions to ask your current LSP about quality

Global businesses rely on LSPs to deliver high-quality translations that resonate with their target audiences. Here are a few questions to ask your current LSP that will help you understand their approach to quality — and, ultimately, mitigate the risk of paying for subpar translations.



Questions to ask your current LSP about quality

How do you define quality?

Having a clear, well-considered definition of what makes a high-quality translation, along with quantifiable criteria, is a must.

How do you measure quality?

Many LSPs have an internal quality assessment program in place. But the details matter. Be sure to ask about the frequency of their audits, how they reduce subjectivity, and the specific framework and error typology they've chosen.

Do you have different quality evaluation processes for human or AI output?

The MQM methodology is applicable to both. But especially if your LSP is not using MQM, you'll want to understand whether and how quality is evaluated across all workflows

How do you evaluate your translators?

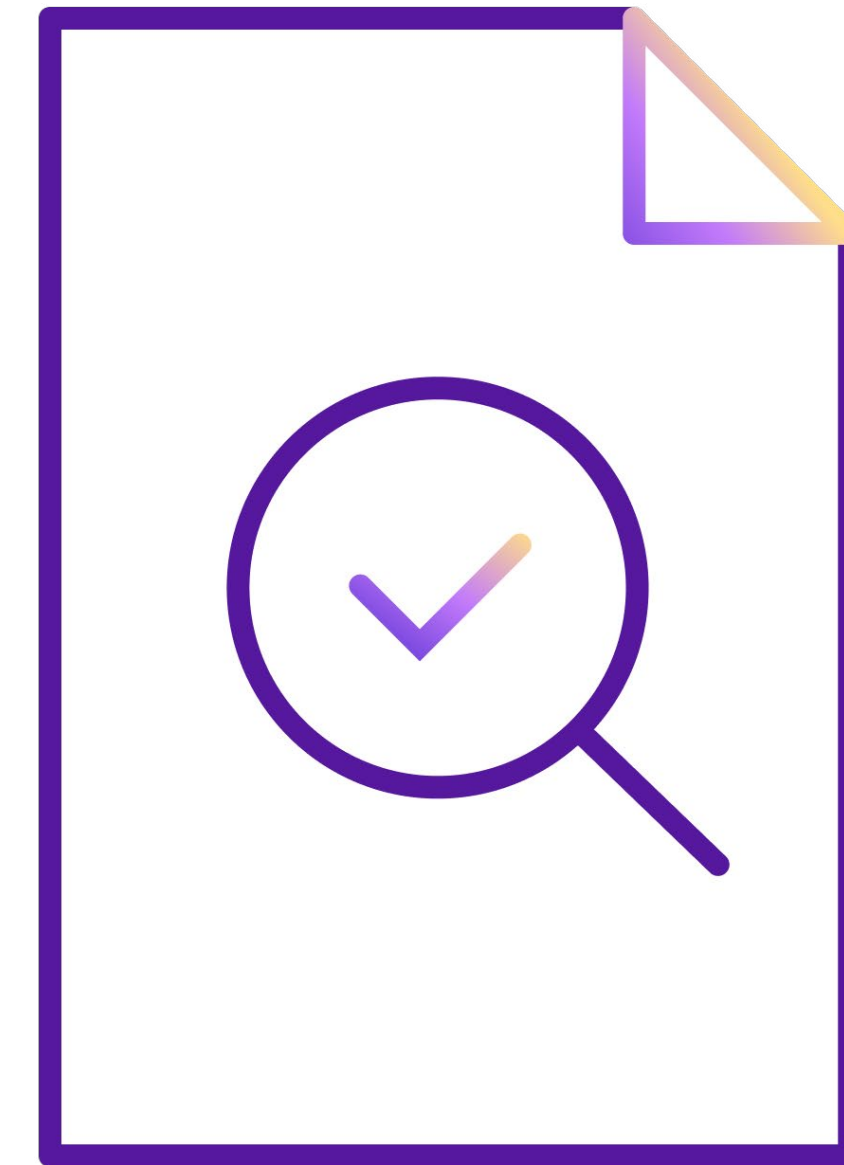
A selective vetting process is crucial. So is having an ongoing evaluation and training program.

What is your process for revising translations if issues are found?

Revisions should be handled promptly. Also consider any fees associated with addressing quality issues and revising the translations.

Conclusion

At Smartling, quality evaluation is not a nice-to-have or a once-in-a-while undertaking but a critical part of the workflow.



Conclusion

That's why we've invested in creating a transparent, comprehensive quality assessment program, complete with regular independent audits. The consistent results prove with confidence that we can guarantee quality. No matter the workflow, we deliver on our promise of high-quality translations time and again.

In addition, Smartling enables companies to be active participants in quality management. Smartling's quality evaluation toolkit, which customers can access without leaving the Smartling platform, is intended to fit into existing workflows seamlessly. Customers can say goodbye to the manual steps associated with external quality evaluation processes while minimizing errors, increasing objectivity, and improving their translations.

Finally, our ongoing investment in developing our LanguageAI platform ensures we remain pioneers of technological innovation in the translation and localization industry. And Smartling customers will continue to reap the benefits.

Are you already translating or thinking about translation to expand your multilingual footprint?

Get in touch to learn how we empower
businesses to create meaningful customer
experiences in every language.

Learn more at smartling.com

About Smartling

Smartling's LanguageAI platform enables enterprises to transform content across devices and mediums. The company is recognized by CSA Research as the No. 1 rated language-oriented TMS, and by users on G2 as the No. 1 rated translation management system. Smartling uses a data-driven approach to localization, which enables its customers to achieve higher-quality translation at a lower total cost.

Smartling is the platform of choice for hundreds of B2B and B2C brands, including IHG Hotels & Resorts, Shopify, Momentive (formerly SurveyMonkey) and Lyft. Smartling is headquartered in New York, with offices in Dublin and London.



hi@smartling.com

1 (866) 707-6278

smartling.com