Harmonising linguistic validation with AI:
Precision, efficiency, and the human touch in patient-reported outcome translation

Helen Williams

The Word Desk, Oxfordshire, UK

Correspondence to:

Helen Williams
info@theworddesk.com

Abstract

The surge in artificial intelligence (AI) prompts a reassessment of linguistic validation methods for patient-reported outcome (PRO) measures. The robust linguistic process, designed to adapt PRO measures for different cultures and languages, is upheld by regulators and the outcomes research community for its value in maintaining concept equivalence across global trial data. Its methods are entrenched in human translation and review, making it more challenging to integrate AI (machine learning, deep learning, natural language processing) compared to other parts of the global localisation industry. This article provides an overview of the key challenges in integrating linguistic validation and AI. Despite these hurdles, it advocates for the industry to embrace the potential benefits through collaborative and responsible innovation.
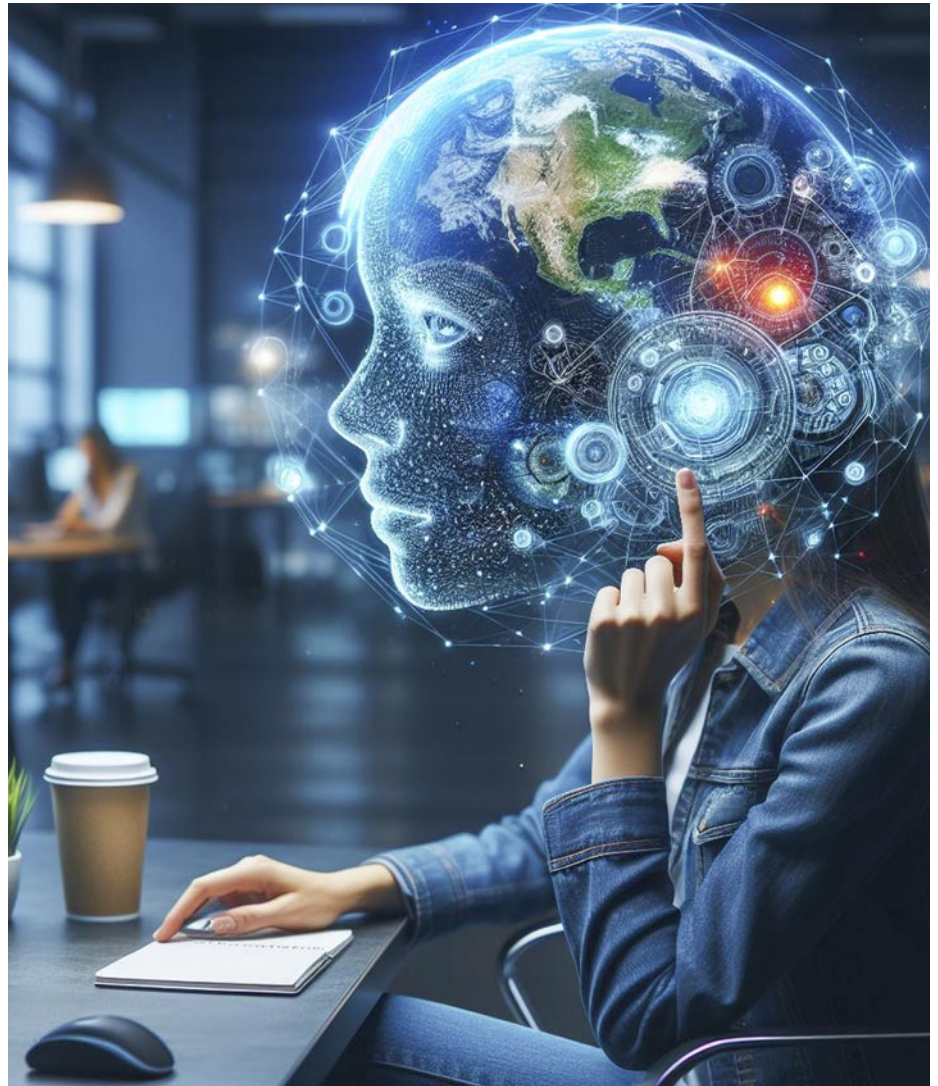
Introduction

Linguistic validation refers to the process of translating a patient-reported outcome (PRO) measure for use in a new language or culture, while ensuring that conceptual, item, semantic, and operational equivalence between instrument versions is preserved.[1] Its objective is to ensure the validity and integrity of patient-reported data gathered across different linguistic and cultural contexts and its significance extends to ethical considerations, regulatory compliance, and the overall improvement of patient-centred outcomes.

Industry standards

The EMA[2] and FDA[3] underlined the need for evidence of PRO translation and linguistic validation in 2005 and 2009 respectively, requesting details of the process used, descriptions of patient testing, language- or culture-specific concerns, rationale for decisions made to create new versions, copies of translated or adapted versions, and evidence that content validity and other measurement properties are comparable between the original and new instruments.

The International Society for Pharmaco-economics and Outcomes Research (ISPOR) Taskforce for Translation and Cultural Adaptation produced best practices for PRO measure translation in 2005.[4] In 2013, The International Society for Quality of Life Research (ISQLR) published recommended standards in a similar vein.[5] Both underscore that qualitative assess-

ment by means of cognitive debriefing interviews with patients from the target population, representing the intended PRO audience, is required for proper translation evaluation. They are now decades-old; nonetheless, these guidelines consistently yield high-quality results when correctly implemented. The methods are grounded in human intelligence, involving a range of highly qualified and trained stakeholders.

Technology in linguistic validation

The existing guidelines lack provision on the integration of translation technology into linguistic validation and it is imperative that industry thought-leaders reassess the process in light of AI advancements, ensuring that the linguistic validation discipline evolves to cope with a rising translation volume and the growing complexity of the digital landscape. In contemporary, patient-centred outcomes research, diverse electronic platforms such as mobile devices, social media, email, SMS, video, and augmented reality solutions will play an increasingly pervasive role. Let us examine how technology currently plays into the typical linguistic validation process.

Figure 1 is an illustration of the basic flow. Variation among language service providers (LSPs) is typically confined to the execution of specific methodological steps. For instance, differences may arise in the choice between single or dual back translation, or the preference for virtual or face-to-face cognitive debriefing. Of vital importance to any project, as in the regulatory guidance, is an audit trail of every linguistic step, edit, and decision made. Of the six steps shown, none are without human input and the humans involved are highly specialised, technical experts. They include ISO-qualified linguists with medical translation experience, trained linguistic validation project managers, clinicians and patients in the target therapeutic area, and instrument developers with advanced degrees in psychology, health outcomes research, or a related discipline. The diverse and specialised nature of these roles, as well as the fundamentally complex and nuanced cultural and conceptual equivalence exercise at play in linguistic validation, underscores why it would be inadvisable to replace their collective expertise with technology, no matter how advanced.

The level of technological input in this workflow differs between LSPs but most use both computer-assisted translation (CAT) and translation management system (TMS) tools. These enhance efficiency and consistency; the former providing translation memory, terminology databases and workflow management functionality, the latter providing collaboration and automation tools to aid the coordination of

> Of vital importance to any project, as in the regulatory guidance, is an audit trail of every linguistic step, edit, and decision made.

Linguistic validation >>>>>>>>>>

1 CONCEPT ELABORATION

PRO measure health-related concepts explained. A roadmap for linguists.

2 ADAPT OR DUAL FORWARD TRANSLATE

Source > target FT

Adapt when "same language, different country"

3 RECONCILE

Best of both version of the two forward translations

4 BACK TRANSLATE & REVIEW

Single or double back translation options and review by LV PMs

Developer/author/clinician review if required

5 COGNITIVE DEBRIEFING

1-1 interviews with target population patients and report discussion between PMs and linguists

6 QA

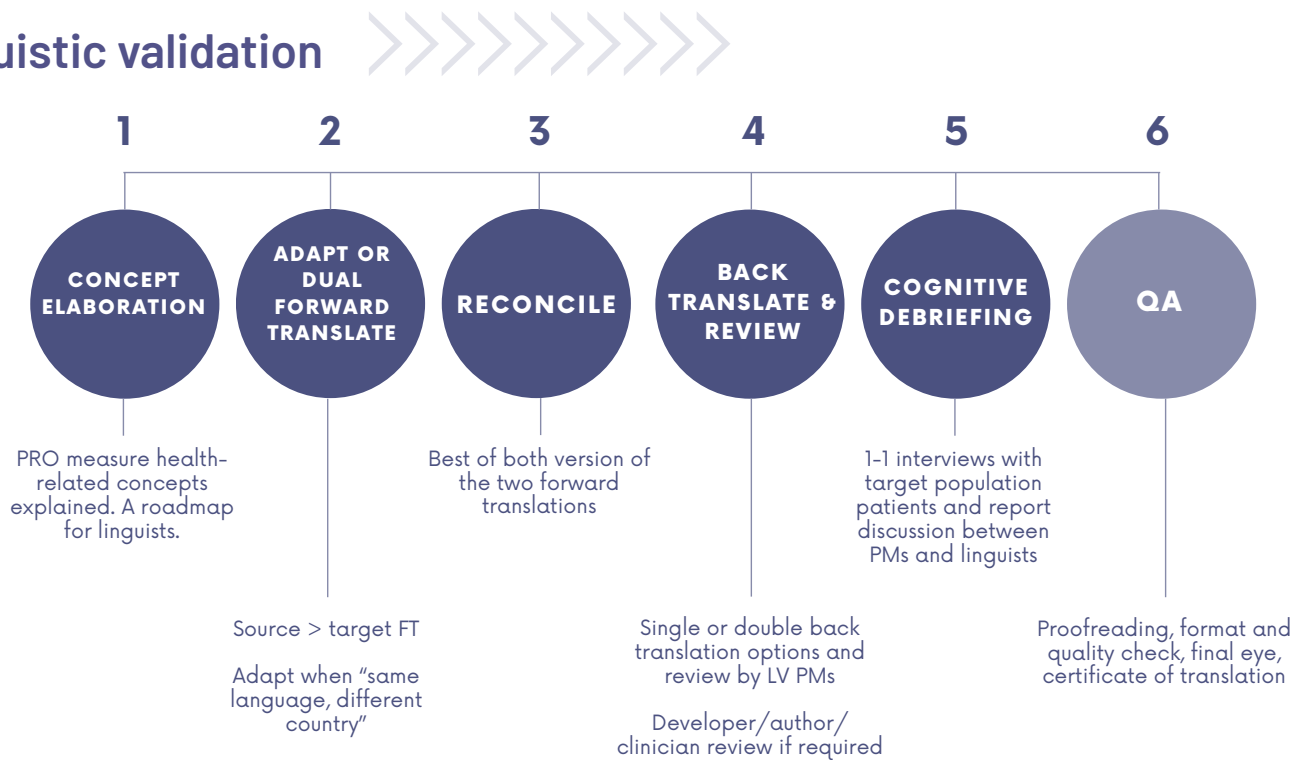Proofreading, format and quality check, final eye, certificate of translation

Figure 1. The Linguistic Validation process from concept elaboration to quality assurance

This figure and article do not address the linguistic validation processes related to multilingual electronic Clinical Outcome Assessment (eCOA) migration or testing and AI, as this would merit another article. Graphic by Helen Williams

Abbreviations: PRO, Patient-reported outcome; FT, Forward translation; LV PMs, Linguistic Validation project managers

tasks among multiple contributors. These aid the human translation process and ease the audit burden by accommodating trackable versioning and reporting.

Leveraging AI: challenges

There is no linguistic validation-specific industry guidance that advises replacing any or all of the six human translation, review, or test steps shown in Figure 1 with wholly automated translation. It is possible, though, that the focus could shift from humans translating the source content, to the humans correcting and augmenting machine-generated translations, or the machine and the human producing a version each for comparison. Ongoing and historic translation content of PROs could be stored to produce PRO-specific memory banks to inform AI machine translation, but this would involve considerable set-up and maintenance work, plus potentially lengthy approval processes. This logical starting point would also be complicated by the complex ownership of PRO measures and the fact that PRO measures are not naturally word-rich or lengthy. They tend to be succinct, short instruments to avoid patient burden, and AI models require vast datasets to perform well (think of the word count of a car manual, for example, in comparison to a 50-word instrument). PRO data is often proprietary and sensitive, owned by the entities conducting the clinical trial. Striking a balance between leveraging PRO translations for AI advancements and respecting ownership rights and ethical standards becomes crucial in the development of AI models in this domain. CAT tools offer access to public term banks for translators; however, in the context of linguistic validation, it is imperative to disable these features. This precaution is taken due to the potential risks associated with uncertain origins of those translations.

How can we leverage AI's capabilities to optimise the linguistic validation process? There is a growing demand to integrate AI in this field, where the predominant clinical study requirements revolve around the classic triad of quality, time, and cost-effectiveness and AI is known to help with efficiency and scalability. Understandably, questions start to mount, as stakeholders see its positive impact in other industries

> There is no linguistic validation-specific industry guidance that advises replacing any or all of the six human translation, review, or test steps with wholly automated translation.

or note that AI is lessening the burden of other life sciences translation tasks. Industries like e-commerce employ neural machine translation (MT) through deep learning models for automated text translation and they can produce a high volume of translated content at speed and relatively low cost compared to human translation. Depending on the model's depth of knowledge, neural MT can also offer contextual translations. The distinguishing factor lies in industries like e-commerce having access to substantial volumes of high-quality language data for training the models. AI including neural MT represents significant risk for the linguistic validation discipline, as it would, for example, for the legal industry, where highly sensitive content is at stake. "Good enough" simply will not do for linguistic validation; this content needs white glove treatment. It's understood that AI can help with speed and scale, but its limitations in terms of accuracy are concerning.

Human review, translation, and testing necessarily underpins linguistic validation, due to the intricate nature of health concepts, cultural subtleties and idiomatic expressions used in patient communities. AI is known to struggle with contextual or cultural understanding, compared to humans. Cultural appropriateness is key to patient understanding, comfort and engagement; if PRO language isolates and confuses the patient, they will neither be inclined nor able to participate in the assessment. They may even find the translations offensive or discriminatory. PRO translations that meet or exceed the industry-recommended minimum standards "will increase the likelihood that the evidence generated […] reliably and validly represents the patients' perspective on health-related outcomes."[5] The stakes are high; if translations are substandard and their measurement properties are negatively affected, it can prevent evidence being used to inform clinical and health policy decision making. The robust coupling of the trusted linguistic validation methodological

framework with tried-and-tested project management best practices produces reliable, auditable results. It is understandable, therefore, that the "unknown quantity" characteristics of AI may alarm industry professionals more than excite them with its potential. We can get the quality output that we need, reliably, so what does it matter if the process is a little clunky and expensive?

Several challenges impede the swift integration of AI into linguistic validation. These include ensuring compliance with regulatory standards, navigating domain-specific nuances and cultural sensitivities that pose challenges in training AI language algorithms, organisational risk aversion stemming from concerns about accuracy, biases and the nuanced understanding required for context. Additionally, issues such as data security, intellectual property, and privacy concerns arise, as AI systems often necessitate access to extensive datasets for training. Furthermore, establishing effective interdisciplinary collaboration between linguistic validation, scientific, and technology teams remains challenging due to prevalent operational silos. Methodologically, the key question revolves around striking the right balance between human expertise and artificial intelligence, aiming for efficient study execution without compromising the integrity of the final deliverables.

Responsible innovation

In the specialised realm of linguistic validation, where precision takes precedence, a discerning approach is vital, favouring judiciousness over a singular pursuit of cost and time efficiency.

> Establishing effective interdisciplinary collaboration between linguistic validation, scientific, and technology teams remains challenging due to prevalent operational silos.

Responsible innovation, guided by collaborative decision-making by qualified industry experts, should steer this transformative journey. As AI's role expands in the broader localisation market, linguistic validation experts can capitalise on AI insights from diverse fields, applying proven techniques and adapting successful strategies without starting from scratch. While embracing AI for efficiency gains and streamlined processes in PRO translation timelines, rigorous testing and integration must coexist with human review and validation. The future of linguistic validation likely involves a synergy of AI, machine translation, and human expertise, promising improved accuracy, enhanced

References

1. Muehlhausen W, Byrom B. Electronic Patient-Reported Outcome Measures. An Implementation Handbook for Clinical Research. Second Edition. 2023:50.
2. EMA. Committee for Medicinal Products for Human Use. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products, 2005. Available from: https://www.ema.europa.eu/en/regulatory-guidance-use-health-related-quality-life-hrql-measures-evaluation-medicinal-products-scientific-guideline
3. US FDA. Patient-reported outcome measures: use in medical product development to support labeling claims. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims
4. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005(2):94–104. doi:10.1111/j.1524-4733.2005.04054.x
5. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013; 22(8): 1889–905. doi:10.1007/s11136-012-0344-y

efficiency, and greater accessibility. Strategies, such as bespoke algorithms and early human reviews of PRO content, address potential challenges, ensuring the training of AI solutions with high-quality language data. The "translatability assurance" step, recommended when a new PRO measure is in development, could extend to assess the content for compatibility with AI. The potential of AI is huge and features such as type-ahead functions, readability highlights, detection of offensive terms, autocorrect, and the capacity to adhere to glossaries and style guides are very attractive. However, in this heavily regulated industry, challenges persist, including AI's limits in the face of contextual understanding, cultural subtleties, depth of domain knowledge, and the risk it presents with potential biases. Human oversight remains indispensable and AI solutions need to be trained within stringent quality control frameworks.

In conclusion, the integration of AI into linguistic validation should not be seen as a threat but as an opportunity for progress. When approached with caution, collaborative decision-making, and a commitment to maintaining high standards, the marriage of AI and linguistic validation holds the promise of more efficient, accurate, and accessible processes. A successful integration, alongside AI algorithms that become increasingly accurate, aware of context, and adaptive, could contribute to accelerated clinical trials, cost reduction in research, and improved data quality. All of these serve our collective goal of helping patients.

Disclaimer
The opinions expressed in this article are the author's own and not necessarily shared by EMWA.

Disclosures and conflicts of interest
The author declares no conflicts of interest.

Author information
Helen Williams, DPhil, is the founder of The Word Desk Ltd, an Oxfordshire-based linguistic validation and eCOA translations agency. Trained initially by experts at Oxford Outcomes and ICON plc, Helen has dedicated the last 14 years to senior leadership roles, scaling language services and businesses within health outcomes research.