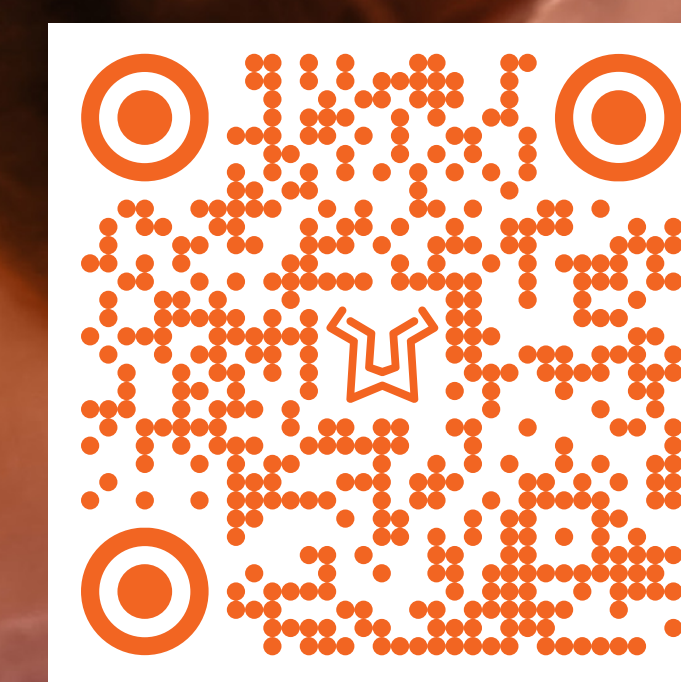


# L'avenir de la validation linguistique est déjà là :

## Améliorer la conformité du processus de réconciliation avec l'IA générative



**LIONBRIDGE** | **Pearson**

Présenté par Lionbridge : Elisabet Sas Olesa ; Karolina Elizondo ; Kathryn Nolte ; Nathalie Azuaje ; Melinda Johnson

## INTRODUCTION

Cette étude se penche sur la manière dont l'intelligence artificielle (IA) générative pourrait soutenir le processus de validation linguistique (VL), en particulier lorsqu'il s'agit d'améliorer l'efficacité de la réconciliation de deux traductions, phase au cours de laquelle deux traductions directes ou plus sont fusionnées en une seule (Koller et al., 2012).

L'objectif principal de notre enquête était d'évaluer la capacité de l'IA générative à détecter les non-conformités dans les réconciliations issues des méthodes conventionnelles.

## ÉVOLUTION DE LA CONCEPTION DE L'INVITE

L'invite initiale a été conçue pour fonctionner en parallèle avec les résultats de la réconciliation traditionnelle, en comparant la Traduction A avec la Traduction B. Elle a ainsi généré un résultat binaire Réussite/Échec concernant la décision finale du réconciliateur.

Cependant, cette approche était synonyme d'importantes limitations, en particulier lorsqu'elle était appliquée à des langues autres que l'anglais et dans des contextes hautement spécialisés, tels que ceux définis par un domaine thérapeutique spécifique ou une maladie fréquemment évaluée dans les instruments de qualité de vie. Face à ces découvertes, nous avons utilisé l'IA générative pour évaluer le résultat de la réconciliation et soutenir ainsi notre processus d'AQ. L'invite mise à jour a ensuite été intégrée

## RÉSULTATS

### Résultats positifs

L'analyse des résultats de l'IA générative par des experts linguistiques a révélé trois domaines de soutien essentiels :

- **Justifications non valides ou incomplètes** : L'IA générative a signalé avec constance les commentaires dans lesquels il manquait une logique linguistique ou conceptuelle suffisante (par ex., « La traduction A est meilleure » ou « La traduction B est préférable »). Elle a également détecté les demandes non résolues ou les questions restées sans réponse. Elle a proposé des recommandations pour ajuster l'invite, en veillant à ce que chaque cas soit remonté pour faire l'objet d'un suivi.
- **Justifications manquantes** : Tous les experts ont noté que l'IA générative identifiait de manière fiable les justifications manquantes, incomplètes ou peu claires, accélérant ainsi fortement la révision de l'AQ.
- **Efficacité** : L'IA générative démontre une vitesse de traitement élevée, en analysant jusqu'à 300 segments de texte en quelques secondes.

Cette présélection permet aux experts linguistiques de filtrer rapidement les résultats et de déterminer si un fichier peut passer à l'étape suivante dans le processus de LV.

## MÉTHODES

Nous avons mené une analyse pratique en utilisant un échantillon PerfO (1 000 à 2 000 mots) qui comprenait des variantes de langues de la région APAC, rares et régionales. Nous nous sommes appuyés sur ces paramètres et avons sélectionné différents fichiers issus d'une localisation précédente de la Clinical Evaluation of Language Fundamentals® - Cinquième édition (CELF-5), un système flexible de tests administrés individuellement. Ces tests aident souvent les cliniciens à diagnostiquer avec précision un trouble du langage chez les enfants et les adolescents (NCS Pearson, Inc., 2013). Les langues cibles sélectionnées pour l'analyse étaient l'espagnol (Argentine), l'espagnol (Espagne), le français (France), l'arménien (Arménie), le japonais (Japon) et le chinois traditionnel (Taïwan).

à *Aurora Clinical Outcomes*, la plateforme propriétaire de Lionbridge pour la validation linguistique de bout en bout. L'utilisation de l'IA générative pour soutenir la prise de décision interne fournit également des informations sur le processus de réflexion du réconciliateur. Elle permet de s'assurer que la tâche a été effectuée dans le respect des normes et exigences du secteur, reflétées dans les OBJECTIFS DE RÉCONCILIATION et mises en avant dans les directives fournies au réconciliateur via les instructions de la tâche :

1

Équivalence conceptuelle de la mesure originale

2

Adaptation culturelle

3

Accessibilité pour la population d'étude/le public cible

4

Détection des tendances aux biais

### Domaines d'amélioration

Malgré le potentiel de l'IA générative, sa nature intrinsèquement imprévisible a également introduit certaines limitations lors de l'exécution de l'invite :

- **Comportement incohérent** : Les résultats de l'IA générative démontrent un comportement incohérent lors de l'évaluation de commentaires identiques entre différents segments, y compris au sein d'un même fichier. Cela compromet la fiabilité du résultat d'IA générative en tant que solution autonome.
- **Difficultés contextuelles/de méta-référence** : L'utilisation fréquente de commentaires vagues et de références telles que « Voir le commentaire ci-dessus » est difficile à gérer pour l'IA générative dans la configuration actuelle. Elle ne parvient pas à trouver le commentaire précédent en question. Les experts suggèrent de préciser, dans les livrables attendus de la part du réconciliateur, qu'il convient d'utiliser des références concises et explicites, ainsi que de répéter les explications dans les segments.
- **Non-standardisation des commentaires de réconciliation viables** : Nos experts s'accordent à dire que la définition et l'alignement sur ce qui constitue un raisonnement et une justification détaillée acceptables doivent être plus précis et explicites entre toutes les parties.

## CONCLUSION

Cette étude démontre le potentiel important de l'IA générative en tant qu'outil d'aide à l'assurance qualité pour identifier les écarts, les non-conformités et les justifications manquantes.

L'IA générative permet de détecter plus rapidement les retouches requises et aide l'équipe de localisation à fournir un retour plus ciblé aux parties prenantes, leur permettant ainsi d'améliorer les performances futures. Cependant, l'expertise humaine reste essentielle pour interpréter les résultats, valider les décisions nuancées et assurer le respect des exigences

réglementaires et liées au projet. Une approche avec intervention humaine (alliant l'analyse par l'IA générative à une révision par un expert) améliore la qualité globale, renforce les performances du réconciliateur grâce à des retours informés, et soutient l'affinage continu du processus.

En définitive, avec une optimisation continue des invites et un entraînement ciblé, cette approche peut fournir des résultats de haute qualité et une diminution des coûts dans les projets de localisation de COA.

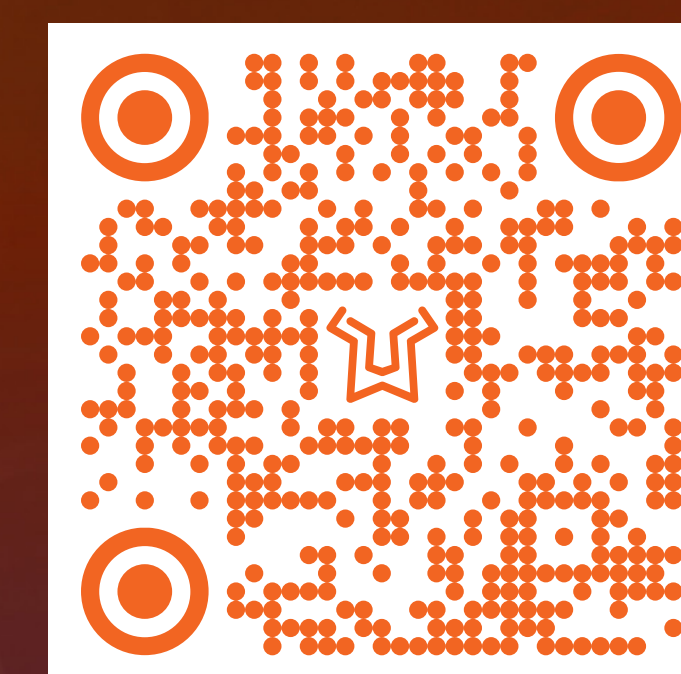
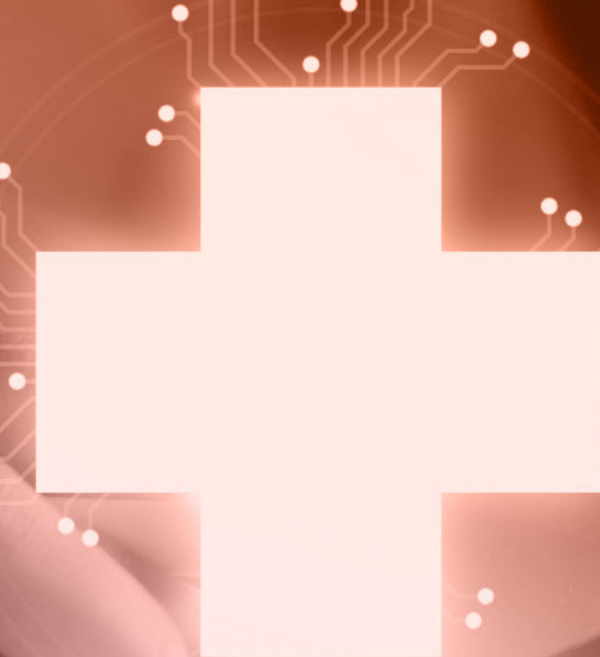
Citations :

Koller, M. et al. (2012). The process of reconciliation: evaluation of guidelines for translating quality-of-life questionnaires. Expert Review of Pharmacoeconomics and Outcomes Research 12(2), 189-197.  
CELF-5 - Clinical Evaluation of Language Fundamentals | Cinquième édition. Copyright © 2013 NCS Pearson, Inc. Tous droits réservés. Web : CELF-5 - Clinical Evaluation of Language Fundamentals | Cinquième édition | Pearson Assessments US



# L'IA générative en action :

## Étude sur l'efficacité des résultats de la révision comparative par l'IA



LIONBRIDGE

Pearson

Présenté par Lionbridge : Stephanie Casale

## OBJECTIFS

La validation linguistique (VL) est le processus qui consiste à localiser et à réviser les évaluations des résultats cliniques, pour une collecte des données exacte et cohérente entre les régions visées. Le processus est long et complexe. Cette approche permet d'assurer la meilleure qualité et les traductions les plus exactes, mais la complexité a un coût. Pour réduire la charge liée à ce processus, notre étude vise à raccourcir les délais et les coûts d'externalisation, tout en respectant les normes élevées exigées.

Cette affiche examine s'il est possible d'utiliser l'IA générative pour réaliser une révision comparative (RC). Il s'agit d'une étape clé de l'assurance qualité dans le processus de VL. Elle consiste à comparer le texte source et le texte rétro-traduit pour déterminer l'équivalence conceptuelle. Étant donné qu'il s'agit d'une étape intermédiaire, les étapes précédentes et suivantes sont effectuées par des linguistes formés et expérimentés, ce qui fait de la RC un candidat idéal pour l'automatisation. Cette approche réduit les risques d'erreurs non détectées avant la finalisation.

Notre recherche visait à développer une invite permettant d'obtenir au minimum la qualité assurée par nos fournisseurs humains actuels pour la révision comparative .

## MÉTHODES

Nous avons d'abord développé une invite produisant la sortie attendue, à savoir un résultat et un commentaire de révision comparative, ce qui apportait davantage de détails sur le résultat. Les résultats de révision comparative se divisaient en trois catégories :

**Identique** – Ce résultat indique que le texte source et la rétro-traduction étaient parfaitement identiques, y compris en matière de majuscules et de ponctuation.

**Équivalent** – Ce résultat indique que, malgré des différences dans la formulation, la structure syntaxique ou d'autres détails, le sens des segments reste équivalent conceptuellement. Un lecteur comprendrait qu'ils transmettent les mêmes informations.

**Doit être révisé** – Ce résultat indique que quelque chose dans les deux segments les rend conceptuellement non équivalents. Un lecteur pourrait ne pas comprendre le texte traduit, en pensant qu'il transmet un élément différent du texte source.

L'invite était ensuite conçue pour produire un commentaire de révision comparative pour tout résultat non identique. Ces commentaires expliquent les différences conceptuelles entre les deux segments, en précisant les erreurs d'interprétation possibles par un lecteur profane. On avait demandé à l'invite d'ignorer toutes les différences de ponctuation et de majuscules, sauf si elles étaient directement liées au sens et à la compréhension. Selon les instructions, l'invite devait également ignorer tout le texte supplémentaire non lié au sens du texte source (c.-à-d. balises de mise en forme, etc.).

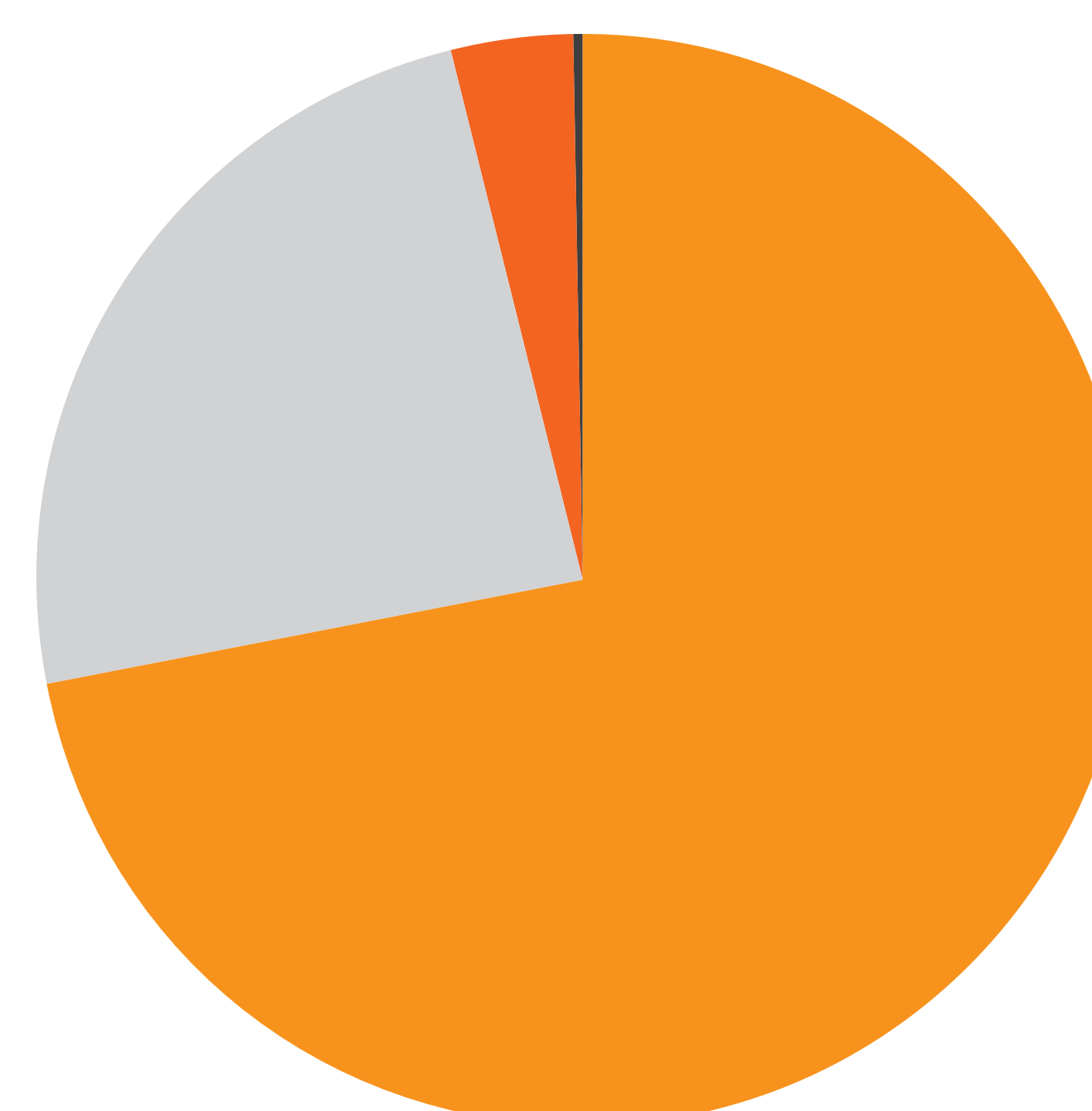
Pour la compilation des données, nous avons utilisé une évaluation Pearson précédemment localisée, Delis-Kaplan Executive Function System (D-KEFS) - Trail Making Test. Elle contient environ 1 000 mots source. Les rétro-traductions ont été effectuées par différents linguistes qui étaient des locuteurs natifs en anglais ou dans la langue cible. Un réviseur comparatif professionnel et des membres de notre équipe de localisation de COA ont mené plusieurs révisions comparatives. Ils avaient différents niveaux d'expérience en révision comparative et différentes langues maternelles, ce qui a permis de comparer un large éventail de résultats.

## RÉSULTATS

Les résultats initiaux sont prometteurs, avec des descriptions concises et claires de l'évaluation originale et des écarts dans la rétro-traduction, le taux de précision préliminaire étant de 96,4 %.

Cette répartition est visible dans le tableau suivant. Il affiche 72,09 % de correspondances exactes entre les résultats des humains et les résultats de l'IA, et indique que l'IA a signalé comme incohérents 24,3 % de segments supplémentaires par rapport au réviseur comparatif humain. On constate que 3,5 % des divergences signalées par le réviseur humain ont été jugées équivalentes par l'IA. Les résultats comprennent également 0,17 % de réponses identiques, ce que les chercheurs ont noté comme risqué en raison du fait que la traduction directe n'était pas en caractères latins et pouvait donc être facilement ignorée si ce texte n'était pas pris en compte.

### POURCENTAGE DE CORRESPONDANCE ENTRE L'IA ET LES RÉSULTATS HUMAINS



- Correspondance exacte 72 %
- IA uniquement 24,3 %
- Humain uniquement 3,5 %
- Risque inhérent à l'IA 0,2 %

#### Pourcentages notables :

Les résultats de l'IA générative analysés par nos experts linguistiques ont révélé trois principaux domaines dans lesquels l'IA générative peut être efficace :

**Risque inhérent à l'IA - 0,17 %** : Ce nombre comprend les segments pouvant avoir été signalés par un humain en raison de l'alphabet non latin dans la traduction directe mais non détectés par l'invite d'IA.

**Réponses incohérentes – 1,26 %** : Au cours de notre révision, l'IA a parfois produit des réponses différentes pour le même ensemble de segments ; cela n'a concerné qu'un peu plus d'1 % des données totales.

## CONCLUSION

**L'IA générative a le potentiel de faire gagner beaucoup de temps et d'argent dans le processus de validation linguistique.**

D'autres études devraient étendre le jeu de données. Les étapes suivantes devront inclure une preuve de concept pour examiner les effets liés à l'utilisation de ce résultat avec des linguistes au cours de l'étape de réconciliation de la révision comparative. Un ajustement supplémentaire de l'invite pourrait aider à atténuer certains risques et incohérences.

