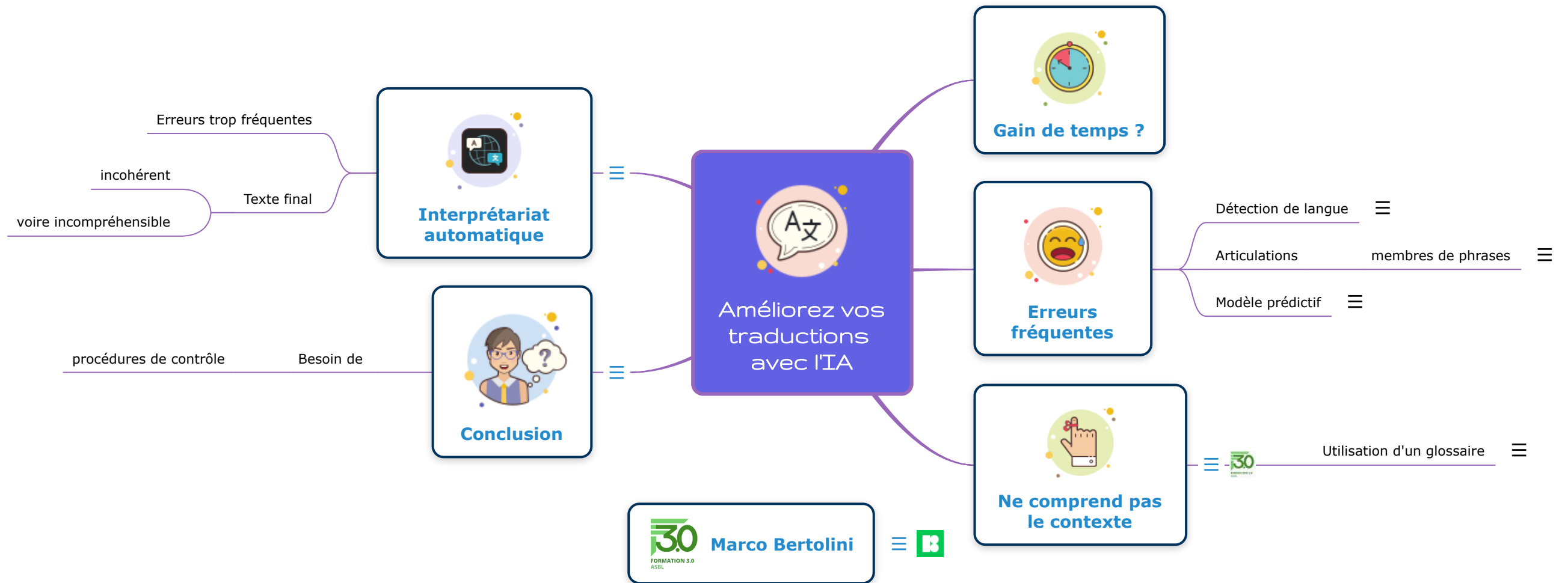




Détection de langue

Un exemple, le verbe “argumenter”. DeepL interprète systématiquement ce verbe comme un substantif norvégien au pluriel. Alors que pour moi, il s’agit du verbe français à l’infinitif. La traduction varie peu : “arguments” au lieu d’“argumenter”. Mais cela casse le sens de la phrase.

membres de phrases

Lorsque je traduis des articles de notre blog dans d’autres langues, les phrases sont découpées en “chaînes de caractères”. Une phrase peut être coupée en plusieurs chaînes, surtout si elle comprend un lien hypertexte. Dans ce cas, parce qu’elle est un modèle linguistique, elle ne peut pas ajuster les chaînes entre elles. Elles sont traitées comme des suites de signaux indépendantes les unes des autres. Par exemple, j’ai traduit récemment un article sur la méthode ABC Learning Design. A un moment, apparait une chaîne qui se termine en français par “je vous ai présenté la”. La chaîne suivante commence par le mot “méthode”. En italien, en espagnol et en portugais, méthode est masculin (metodo). La chaîne “ je vous ai présenté la” devient “vi ho presentato la” en italien, mais est suivie de “metodo” qui est masculin. Cela veut dire que je dois corriger chaque chaîne contenant les articles féminins et le substantif masculin. Ce n’est pas dramatique, mais cela diminuerait notre crédibilité chez les lecteurs italiens, espagnols et portugais.

Modèle prédictif

Les grands modèles linguistiques sont des systèmes probabilistes. Quand ils rencontrent une chaîne de caractères, ils évaluent le pourcentage de la suite. Par exemple, dans une phrase qui dit “la population française en”, le modèle calculera qu’il y a 90% de chances pour que le mot suivant soit “général”. Dans certaines chaînes de caractères, la probabilité est si élevée que le modèle ajoute le mot “général” de sa propre initiative. Alors que, comme mon contexte est académique et professionnel, le mot suivant est probablement différent. Mais si je ne relis pas cette chaîne, je vais me retrouver avec une traduction italienne ou portugaise qui contiendra le mot “général” alors qu’il n’apparaît pas dans la phrase originale et n’a aucun sens ni aucune fonction dans la traduction attendue.

Ne comprend pas le contexte

Un même mot peut avoir plusieurs sens, différents selon le contexte. Par exemple, j’utilise le mot “qualification” dans un contexte professionnel. Il signifie donc “aptitude, ensemble de compétences liées à un métier, une profession”. Mais DeepL traduit systématiquement “qualification” par “sélection” car ce mot signifie “être classé, être sélectionné” dans le monde du sport. Heureusement, une fonctionnalité permet de “fixer” le type de traduction souhaité. Mais, cet exemple, comme d’autres montre bien qu’il faut constamment relire et vérifier les traductions. Même s’il y a gain de temps, cela ne signifie pas qu’il ne faille pas faire réviser par une personne. C’est pourquoi la législation européenne sur la transparence de l’IA dispense les contenus révisés par un humain de l’étiquetage des contenus gérés par l’IA générative.

Utilisation d'un glossaire

Dans DeepL aussi bien que dans Translate Press, il existe une fonctionnalité qui permet de forcer le logiciel à traduire le mot d'une certaine façon. C'est le "glossaire". C'est une sorte de dictionnaire ou de thésaurus qui contient une liste de mots et la traduction retenue. Par exemple, quand nous avons créé le MOOC Dys, nous expliquions dans la première partie du cours que les troubles d'apprentissage, tels que la dyslexie, la dysphasie, etc. étaient des difficultés issues du développement cérébral, mais pas des maladies. Or, à plusieurs reprises, le logiciel a traduit "troubles" par "maladie". Ce que des participants du MOOC nous ont rapidement signalé. Avec la fonctionnalité "glossaire", nous avons pu revoir l'ensemble des traductions en nous assurant que cette erreur ne se reproduirait pas.

Conclusion

Les erreurs signalées ci-dessus ne sont pas dramatiques en soi. Mais, elles soulignent le fait que toute création générée par l'intelligence artificielle générative devrait faire l'objet d'une révision humaine. Et que, quand vous traduisez un texte dans une autre langue à l'aide de l'IA il vaut mieux avoir une connaissance de base de cette langue et de ses mécanismes. Sous peine de perdre toute crédibilité auprès de votre public.

Interprétariat automatique

YouTube a récemment intégré une traduction instantanée automatisée, basée sur un LLM. Les résultats sont souvent hilarants. Des mots sont mal traduits, voire carrément inventés par le modèle. A tel point que la vidéo originale en devient pratiquement incompréhensible. Prions pour que les organisations internationales n'adoptent pas ce type de système pour régler les conflits entre états...

Marco Bertolini

Carte mentale publiée sous licence Creative Commons International 4.0 BY-SA.

Les icônes utilisées dans cette carte proviennent du site Icon8 [Rappel](#) icon by [Icons8](#)