

Résumé français

Née dans l'immédiat après-guerre, la Traduction Automatique a drainé des ressources considérables, et a impulsé un formidable élan en termes de Recherche & Développement en Traitement Automatique des Langues, en linguistique de corpus, tout comme en linguistique formelle. À l'heure où les grands acteurs du web nous promettent de la traduction « speech-to-speech » abordable et en temps réel sur nos terminaux mobiles (smartphones, tablettes et autres objets connectés), quel bilan peut-on dresser de ce qui apparaît, au départ, comme une utopie, vieille de plus de 50 ans : traduire, sans comprendre le sens, le contexte culturel, l'implicite, en bref, la Valeur des unités linguistiques ? Après avoir replacé la naissance de la Traduction Automatique dans son contexte historique, culturel et technique, nous examinerons d'un œil critique les limites mais également les succès enregistrés ces dernières années. Nous dresserons un bilan critique de la TA guidée par les données, en cherchant à dépasser les seules considérations techniques. Les approches actuelles mettent en avant la prédictibilité des patrons linguistiques (unités « préfabriquées »), tout en étant limitées à la fois par la quantité et la qualité des corpus de traduction disponibles, ainsi que par une incapacité structurelle à la généralisation. Nous défendrons l'idée que les succès (relatifs) des approches statistiques nous amènent, malgré tout, à repenser fondamentalement notre rapport à la langue.

Abstract

Machine Translation is often seen as one of the achievements that occurred just after WWII. Since the 1950's, it has drained significant financial resources, but it also led to a major boost in Research & Development in the fields of Natural Language Processing, corpus linguistics, as well as formal linguistics. At a time when Tech giants are promoting real-time "speech-to-speech" translation on our mobile devices (smartphones, tablets and other connected objects), what evaluation can be made of this apparent utopia 50 years down the track? Can translation be achieved without grasping the meaning of individual linguistic units and the cultural context in which they are used? What about the implicit dimension? In short, can machines translate without the "meaning" of linguistic units?

We will start this critical assessment by first setting MT in its historical, cultural and technical contexts. We will examine the limitations of MT, but also the unexpected achievements recorded in recent years, while trying to go beyond the purely technical considerations. Contemporary approaches emphasise the predictability of linguistic patterns ("prefab" units), while being heavily limited by the quantity and quality of available translation corpora, and, at the same time, they appear structurally inadequate for the creation of linguistic rules from one-off cases. We will argue that the (relative) success of statistical MT leads us, nonetheless, to reconsider our relationship to language as a whole.

Hanna Martikainen, *Sources de biais traductionnel dans la traduction humaine et la traduction automatique post-éditée de résumés médicaux Cochrane*

Résumé français

L'objectif de cette étude est de comparer la traduction humaine (TH) et la post-édition de traduction automatique (TA) dans le domaine médical en termes de biais traductionnel. Les revues systématiques Cochrane font référence en matière de médecine factuelle, et la traduction de leurs résumés joue un rôle important dans la transmission multilingue du savoir médical. Cochrane France, qui assure depuis 2010 la traduction en français des résumés Cochrane, a d'abord fait appel à des agences de traduction spécialisées, avant d'opter en 2013 pour une solution de TA statistique. La présentation des résultats de la recherche doit idéalement être exempte d'erreurs et aussi objective que possible. Ainsi, nous définissons les sources de biais traductionnel comme des éléments dans ces textes traduits susceptibles d'interférer avec cet objectif communicatif premier et d'influencer l'interprétation des caractéristiques des essais étudiés (interventions, participants, critères d'évaluation etc.), de leurs résultats, ou de la confiance des auteurs en ceux-ci. Nos premières observations sur corpus suggèrent de potentielles différences entre les traductions produites par ces deux processus en termes de types d'erreurs et de traductions biaisées. Tandis que des erreurs de terminologie et de syntaxe sont observées dans les deux sous-corpus, de telles erreurs apparaissent à un niveau plus élémentaire dans les textes post-édités, probablement en raison de l'influence de la TA. Le sous-corpus de TH, en revanche, comporte davantage de traductions biaisées de marqueurs modaux, en particulier d'atténuateurs, ce que nous imputons au degré plus important d'interprétation qu'implique la traduction classique.

Abstract

The aim of this study is to compare human translation (HT) and post-edited machine translation (MT) in the medical field focusing on translational bias. Cochrane systematic reviews provide up-to-date, impartial information on the effectiveness of interventions, in order to support evidence-based medical decision-making. Translating Cochrane review abstracts plays an important role in the multilingual diffusion of medical knowledge. Cochrane France has been translating Cochrane abstracts into French since 2010, first using specialised translation service providers and since 2013, by integrating statistical MT and partly volunteer post-editing into the process. The presentation of research results should be exempt from errors and as objective as possible. We define sources of translational bias,

then, as elements in these translated texts that are likely to interfere with this communicative objective, and have the potential to impact on the readers' interpretation of the characteristics of the test cases (interventions, participants, outcomes, evaluation criteria), their results, or the authors' confidence in them. We will present here some first corpus observations from this study in an attempt to characterise the potential differences between translations produced by the two processes. Our preliminary research suggests that HT and post-edited MT differ in terms of errors and biased translations. While terminology and syntax errors are observed in both sub-corpora, such errors appear at a more elementary level in the post-edited texts, most likely due to the influence of MT. The HT sub-corpus, on the other hand, contains more biased translations of modal markers, especially "attenuators", which we attribute to the greater degree of interpretation involved in traditional translation.

Thierry Poibeau, *Quelle qualité pour la traduction automatique ?*

Résumé français

La traduction automatique est en train de connaître une nouvelle révolution avec l'avènement des approches neuronales. Si la base de ces systèmes reste statistique, les approches sont maintenant fondées sur l'analyse de la phrase globalement et ne nécessitent plus d'assembler des morceaux disparates de traduction, comme le faisaient les approches statistiques jusqu'ici. De réels progrès sont déjà visibles avec ces approches. Quelle qualité peut-on espérer in fine ? Se rapproche-t-on ainsi de l'approche humaine ? Quelles limites sont à prévoir ? Mon exposé donnera quelques pistes autour de ces questions.

Abstract

Machine translation is deeply renewed these days thanks to neuronal approaches (aka 'deep learning' approaches). Although these systems can still be called 'statistical', they are very different from previous ones since they process at sentence level and no longer need to generate sentences from scattered pieces of translation. How far can we go with this approach? To what extent can it be compared to human translation? What are the inherent limitations of this approach? We will provide a few potential solutions as we answer these questions during the talk.

Rodolfo Maslias, *In Termino Qualitas*

Résumé français

Le défi de produire de la législation en 24 langues, qui fait partie du droit national de 28 pays, et le rôle de la terminologie pour la consistance linguistique et juridique de ces textes, qui ne sont plus des traductions mais des originaux faisant foi. La coopération entre les institutions de l'UE en ligne en terminologie via le portail EurTerm qui constitue à la fois une plateforme de communication et un outil collaboratif qui permet le travail interactif pour les projets terminologiques de toutes les institutions et agences de l'UE. La collaboration au niveau de chaque langue (par des wikis impliquant des experts nationaux). Les ressources terminologiques mises à disposition du public via IATE-public et la page web publique termcoord.eu. Les nombreux projets IATE avec des universités permettant aux professeurs et étudiants de travailler en utilisant les mêmes méthodes et outils que les traducteurs et terminologues des institutions européennes. La terminologie proactive consistant à produire au préalable de la terminologie dans toutes les langues officielles en relation avec chaque texte à traduire en utilisant des extracteurs de termes. L'interopérabilité des outils avec la création des APIs pour la nouvelle version d'IATE et les services web pour sa connexion avec d'autres bases de données. Enfin, les méthodes techniques d'insertion de terminologie dans les outils de traduction assistée (bases de données extraites de IATE via TBX, reconnaissance de termes etc.) ainsi que le rôle de la terminologie dans la traduction automatique ex ante, en tant que ressource (seulement pour la terminologie normative) et en tant qu'outil de contrôle de qualité, ex post.

Abstract

The challenge of creating of legislation in 24 languages as part of the national legislation of 28 countries; the role of terminology for the linguistic and legal consistency of acts that are not translations anymore but texts with the same validity as the originals. The online cooperation between the institutions of the EU in terminology via the portal EurTerm that constitutes, at the same time, a communication platform and a collaborative tool permitting the interaction for the terminology projects of all EU institutions and agencies. The collaboration at language level via wikis that also involve national experts. The terminology resources that are made available to the public through the public webpage termcoord.eu. The numerous IATE projects with universities allowing professors and students to work with the same methods and tools as the translators and terminologists of the EU institutions. The proactive terminology that involves preparing in advance terminology in all the official languages for the texts sent to translation. This is done using of term extractors. The interoperability of the tools with the creation of APIs for the new version of IATE (under construction) and with the web services connecting it to other databases. Finally, the technical methods for inserting terminology into the CAT tools (databases extracted from IATE via TBX, term

recognition etc.) and the role of terminology in machine translation *ex ante* as a resource (normative terminology) or *ex post* as a tool for quality control.

Julia Caratini, *Utilisation des outils de TAO dans le milieu professionnel*

Résumé français

Présentation sur l'utilisation des outils de TAO (Trados) au quotidien dans le milieu professionnel, l'utilisation de la traduction automatique et le développement de la post-edition.

Abstract

Presenting and using CAT tools in a professional setting, the use of machine translation and the development of post-editing.

Nadège Do Carvalho, *Intégration de la TA dans SDL Trados Studio : démonstration*

Résumé français

Intégration de la TA dans SDL Trados Studio : démonstration

Abstract

How to integrate MT in SDL Studio

Jean-Christophe Bailbe, *Enseigner la TAO : une expérience*

Résumé français

L'ère est numérique et les « nouvelles » technologies ont investi ou investissent la plupart des secteurs, notamment celui de la traduction. De nombreux outils sont à la disposition des traducteurs. Les connaître, parvenir à en maîtriser certains, est désormais une nécessité plus qu'un luxe. Dans ce contexte, il est intéressant de se pencher sur l'approche et l'enseignement de l'usage de ces outils, sans perdre de vue la finalité principale de l'activité de traduction : transposer le sens d'un message d'une langue source vers une langue cible. Lors de mon intervention, je tenterai de faire part de mon expérience d'utilisateur, de formateur et de chargé de cours en TAO.

Abstract

The "new" technologies of the digital age pervade most sectors, including the translation industry. A large number of tools are available to the translators. Knowing these tools and having a reasonable command of them is now a must. In this context, we can reflect on teaching how to use these tools, without forgetting the very purpose of translation: conveying the meaning of a message from a source language to a target language. In my presentation, I will try to describe my experience as a CAT tool user, trainer and teacher.

John Di Rico, *Wordfast versus the Machine: Demonstration*

Abstract

This presentation will show the main features of Wordfast Pro 4. A workflow approach from translation to delivery will be used and the presentation will focus on features that improve consistency, facilitate communication between translators and editors, and speed up the overall translation process.

Résumé français

Cette session présentera les principales capacités de Wordfast Pro 4. Une approche montrant le flux de travail allant de la traduction à la livraison sera utilisée et la présentation portera sur les fonctionnalités qui améliorent la cohérence, facilitent la communication entre les traducteurs et les éditeurs et accélèrent le processus de traduction en général.