

Tralogy II. Synthèse de la session 2

Holger Schwenk

► **To cite this version:**

Holger Schwenk. Tralogy II. Synthèse de la session 2. Tralogy II. Trouver le sens : où sont nos manques et nos besoins respectifs ?, Jan 2013, Paris, France. 3p. hal-02497271

HAL Id: hal-02497271

<https://hal.archives-ouvertes.fr/hal-02497271>

Submitted on 3 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TRALOGY

TRALOGY II. **Synthèse de la session 2**

Holger Schwenk

TRALOGY II - Session 2
Date d'intervention : 17/01/2013



La deuxième session de la conférence Tralogy 2013 a abordé l'importance et l'utilisation du sens en traduction automatique. Lorsque l'on traduit une phrase d'une langue dans une autre, le but principal est bien entendu de préserver le sens de la phrase initiale. Il se pose alors la problématique très intéressante de savoir si une machine pourra réellement comprendre le sens d'une phrase comme un être humain et choisir la bonne traduction, parmi un grand choix de traductions possibles ?

La session a comporté trois exposés.

La première intervention invitée de David Farwell est intitulée « *Pragmatics and High Quality MT* ». Il a montré à l'aide d'un exemple très intuitif que le contexte social est très important lors de la traduction d'un texte : comment traduire l'expression « *the ground level of a building* » ? Ceci semble pourtant très simple, la grammaire de l'expression est très simple et les mots ne semblent pas être ambigus. Cependant, la bonne traduction nécessite une connaissance de son contexte. En effet, en Europe « *the ground level* » d'une maison à plusieurs étages correspond au niveau de la rue, alors qu'aux États-Unis la même expression fait référence au premier étage. Une traduction devant préserver le sens d'une phrase, il faudrait donc adapter la traduction au contexte régional. D'autre part, la traduction exacte de l'étage est primordiale lorsqu'il s'agit par exemple d'une description d'un appartement à vendre, mais ceci n'intéresse probablement pas le lecteur dans un article d'actualités. David Farwell argumente donc que la prise en compte du sens et du contexte d'un texte ne dépend pas de la langue cible, mais plutôt de l'audience cible prévue. L'exposé conclut par une comparaison de trois traductions d'une centaine de textes d'actualités de l'espagnol vers l'anglais. Trois catégories d'erreurs ont été dégagées : des erreurs (intentionnelles) de traduction, différentes manières d'exprimer le même sens et finalement des assomptions différentes des traducteurs sur le texte source ou l'audience. 41 % des différences entre les textes concernent la dernière catégorie !

Dans sa communication invitée, Marianna Apidianaki du LIMSI-CNRS à Orsay (France) a présenté des recherches sur l'ambiguïté lexicale en traduction automatique (*Lexical ambiguity in Machine Translation*). Dans la plupart des langues, il y a des mots qui peuvent avoir plusieurs sens, comme par exemple le mot français « avocat ». Il peut signifier le fruit ou un auxiliaire de justice qui représente des personnes en justice. Souvent ces multiples sens correspondent à des traductions différentes dans une autre langue : ici « *avocado* » et « *lawyer* » respectivement en anglais. Notons que les êtres humains sont généralement capables de détecter avec facilité le bon sens du mot grâce au contexte du mot, et surtout, à l'aide de leur connaissance du monde. Cependant, la désambiguïsation lexicale des mots, en anglais *Word Sense Disambiguation (WSD)*, reste un domaine de recherche compliqué. Dans son exposé, Marianna Apidianaki a donné un aperçu des recherches en cours et de leurs applications dont la traduction automatique. Les liens entre la TA et le WSD sont en fait multiples. D'une part, la détection des sens des mots devrait permettre de mieux choisir la traduction appropriée, et d'autre part, le WSD peut exploiter les ressources collectées pour la TA, notamment les textes parallèles. En fait, les traductions différentes d'un mot dans une autre langue peuvent permettre de détecter les sens du mot dans la langue source. Marianna Apidianaki conclut son exposé par une analyse des métriques d'évaluation automatique en traduction : le score BLEU, toujours fréquemment utilisé, ne semble pas permettre de dégager des systèmes qui intègrent des mécanismes de désambiguïsation lexicale, alors que d'autres métriques telles que METEOR ou TER-plus semblent être plus appropriées.

Dans le dernier exposé, Octavian Popescu et Elisabetta Jezek de la Fondation Bruno Kessler (FBK) à Trento en Italie abordent la traduction de segments de mots. Il est bien connu qu'une traduction mot-par-mot ne donne généralement pas de bons résultats, en particulier dans le cas d'expressions idiomatiques. Un exemple très simple est la traduction de l'expression « *How are you ?* » qui doit être traduite en « Comment ça va ? ». Octavian Popescu argumente que les mots peuvent changer leur sens en fonction du contexte, mais que la plupart des segments de mots ont un sens stable, indépendamment du contexte dans lequel ils apparaissent. On peut notamment constater que le sens d'un verbe dans un segment de mots dépend essentiellement

de l'objet direct qui suit. Ceci a été nommé « *sense stable phrase* ». L'utilisation de cette propriété dans un système de traduction de l'anglais vers l'italien permet d'obtenir des traductions qui sont considérées systématiquement meilleures que celles produites par le système de traduction de Google.

La discussion qui a suivi les présentations a clairement montré un grand intérêt des développeurs et utilisateurs de systèmes de traduction automatique à intégrer plus de connaissances sémantiques dans leurs approches. Entre autre, des exemples de systèmes de traduction du français vers l'arabe ont été présentés. La discussion a également dégagé des similarités entre les recherches en traitement du langage naturel et la linguistique de corpus, et il y eut un large consensus sur le fait qu'un renforcement des coopérations entre ces disciplines serait très bénéfique pour tous les acteurs. Thibault Grouas, de la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF, Ministère français de la Culture) a indiqué des travaux sur l'extraction de connaissances des ressources Wikipédia et Wiktionary, proposant de les mettre à disposition de la communauté des chercheurs en traduction automatique.