
TD5 - PRACTICAL ASSIGNMENT 5

Vincent Matthys
vincent.matthys@ens-paris-saclay.fr

Dans le cadre de la compréhension de l'état de l'art en traduction automatique, on a testé [Google Translate \(GT\)](#) et [DeepL \(DL\)](#).

1 Phrase complexe

“There was Eru, the One, who in Arda is called Iluvatar; and he made first the Ainur, the Holy Ones, that were the offspring of his thought, and they were with him before aught else was made.”

Voici la première phrase du *Silmarillion*, roman de J. R. R. Tolkien écrit en anglais et traduite :

- par GT avec le point final - *Il y avait Eru, l'Un, qui à Arda s'appelle Iluvatar; et il fit d'abord les Ainur, les saints, qui étaient la progéniture de sa pensée, et ils étaient avec lui avant que tout autre ait été fait.*
- par GT sans le point final - *Il y avait Eru, l'Un, qui à Arda s'appelle Iluvatar; et il a fait d'abord les Ainur, les Saints, qui étaient la progéniture de sa pensée, et ils étaient avec lui avant tout autre chose a été fait*
- par DL avec et sans point final - *Il y avait Eru, Celui qui à Arda s'appelait Iluvatar; et il fit d'abord l'Ainur, les Saints, qui étaient les descendants de sa pensée, et ils étaient avec lui avant que rien d'autre ne fût fait.*

On remarque d'une part l'importance du point final dans la traduction émise par GT, au niveau des temps des verbes choisis, et dans la syntaxe après la proposition coordonnée introduite par “and” indiquant que le PoS de “before” a changé en cours entre les deux traductions, et indiquant probablement que le parsing et le part-of-speech tagging sont conjointement effectués. On remarque que les noms propres sont correctement laissés à l'identique pour les deux systèmes, même si le choix paraît plus littéral pour “the One” de la part de GT. A l'inverse, DL n'a pas associé “the Ainur” au pluriel de la proposition juxtaposée s'y rapportant, délaissant ainsi le pluriel du déterminant “the”. On note aussi qu'étrangement “the Holy Ones” perd sa majuscule lorsque le point final est présent dans la traduction de GT, pour une raison inconnue.

2 Influence du contexte

“Il aime bien le mouton. Surtout les côtelettes de mouton.” est traduit par DL et GL en *He likes sheep. Especially the lamb chops..* Pour les deux systèmes, le contexte n'influe pas sur le choix de la traduction de “mouton” malgré la référence à de la nourriture, quelle

soit antérieure ou postérieure. De façon similaire, “A mouse appeared. It looked hungry” est traduit par les deux systèmes en *Une souris est apparue. Il avait l’air affamé.*, et le contexte ne permet pas de choisir le genre correct du pronom traduit de “it”. Même en transformant les deux phrases en les reliant par une conjonction de coordination le résultat reste identique.

3 Expressions idiomatiques

“Don’t judge a book by its cover” est traduit par GT et DL en *Ne jugez pas un livre par sa couverture.*, en ignorant donc l’expression idiomatique anglaise. En revanche, “Avoir d’autres chats à fouetter” est reconnu par DL comme une expression idiomatique, et une translation correcte *Have other fish to fry.* est proposée, en précisant l’équivalence des deux expressions idiomatiques, Une expression littérale est également proposée *Have other cats to whip.*. C’est une méthode qui permet de proposer la traduction littérale et l’expression idiomatique lorsqu’elle est détectée.

4 Influence de la paire de langues

La question “What is the influence of a language pair?” a été traduite en français puis en polonais et enfin re-traduite en anglais, on obtient :

- pour GT : *How does the language pair affect?*. Ce n’est pas identique à la question de départ, même si le sens reste conservé.
- pour DL : *What is the influence of a language pair?* qui est exactement la question initiale.

Malheureusement DL ne propose pas encore de langage “rare”, La même traduction en remplaçant le polonais par le sesotho (environ 5 millions de locuteurs, une des langues avec le moins de locuteurs proposée par GT) produit, selon GT *What is the second language influence?*, qui est plus éloigné de l’originale qu’en passant par le polonais.

5 Syntaxe des traductions

Comme montré, le genre de pronom neutre, le temps des verbes, et le PoS des mots peuvent varier entre la traduction automatique et celle attendue, et entre GT et DL. Ce qui indique des méthodes de parsing et de PoS différents entre ces deux systèmes. On peut par exemple mettre en exergue la différence de choix de temps dans la traduction de “It’s bad that i’m going out with him” :

- par GT : *C’est mauvais que je sors avec lui*
- par DL : *C’est mal que je sorte avec lui.*

GT n’utilise pas la construction de avis + proposition au subjonctif pour cet exemple, alors que pour un autre, “It’s too bad it’s raining.”, dont la traduction par GT et DL est *C’est dommage qu’il pleuve*, avec le subjonctif. La traduction de GT semble donc plus littérale que celle de DL qui semble dégager les règles syntaxiques des langues cibles.