



IFT6285 (TALN) — Devoir9
Évaluation de traductions

Contact :
Philippe Langlais +1 514 343 61 11 ext: 47494
RALI/DIRO felipe@iro.umontreal.ca
Université de Montréal <http://www.iro.umontreal.ca/~felipe/>

■ dernière compilation : 24 novembre 2020 (15:31)

Contexte

La traduction automatique a fait grâce à l'apprentissage profond un bon spectaculaire en qualité, augmentant les perspectives d'intégration de cette technologie dans les chaînes de traitement multilingues. Il existe plusieurs systèmes de traduction en ligne de très bonne qualité, donc voici quelques représentants : [DeepL](#), [Google](#), [Bing](#), [Systran](#), [Reverso](#). Les évaluations de tels systèmes sont fréquentes sur les forums du Web, voir par exemple [ici](#) et [là](#). Le but de ce devoir est d'évaluer une ou plusieurs de ces technologies et de recenser des problèmes particuliers que vous observez. Votre évaluation portera sur la paire de langue anglais-français (en traduisant depuis ou vers le français).

À faire

1. Considérez des formes idiomatiques, comme celles listées [ici](#). Soumettez-les à un ou plusieurs traducteurs et concluez à leur bonne ou mauvaise traduction. Par exemple, Google produit la traduction *I have been sitting too long, my legs are ants.* pour la phrase *Je suis resté assis trop longtemps, j'ai des fourmis dans les jambes.*, ce qui est évidemment une erreur.
2. Ce [quiz](#) contient 100 exemples formés d'une phrase source et de deux traductions, l'une produite automatiquement, l'autre par un traducteur humain. Vous devez indiquer sur la ligne **ANSWER** de chaque exemple si la première traduction est selon vous humaine (HUM) ou automatique (NMT). Vous pouvez également répondre ??? pour indiquer que vous n'en avez aucune idée. Remettez votre fichier annoté. Si vous ne lisez pas le français, produisez un tel fichier à l'aide d'un bitexte existant, d'un système de traduction puis annotez le comme demandé (sans bien sûr prendre connaissance de l'origine de la traduction) et remettez également la référence.
3. Vous devez trouver au moins 20 traductions fautives produites par un système de traduction. Vous étiquetterez ces problèmes à l'aide d'un jeu d'étiquettes de votre choix que vous décrirez. Ce jeu peut être simple comme "pas grave / grave" ou plus précis comme faute

d'accord / contresens / mauvais choix lexical / typographie / etc. Le but n'est pas ici de piéger un système, mais d'en mesurer l'usage dans une situation "réelle". Vous remettrez un fichier **erreur.txt** comprenant les phrases sources traduites, leur traduction problématique et une description des problèmes.

Vous pouvez utiliser n'importe quelles données (par exemple le texte de ce devoir). Vous prendrez soin d'identifier la provenance des phrases que vous avez soumises à la traduction, le nombre de traductions évaluées/analysées et le nombre total de traductions avec un problème.

Je vous donne à toute fin utile accès à un fichier de traductions produites par [Google](#) et par [DeepL](#) pour des phrases en français extraites de nouvelles (le type de texte sur lesquels les systèmes sont typiquement entraînés). Le format en est le suivant : une phrase source (SRC), une traduction humaine (HUM), une traduction automatique (dans l'exemple produite par DeepL), suivi de deux lignes indiquant la prédiction de 18 classificateurs entraînés à reconnaître les traductions humaines de traductions produites par un système neuronal (XLM), la classe 0 indiquant une traduction humaine, une prédiction de 1 indique au contraire une traduction machine. La première de ces deux lignes (**Label 0**) indique la prédiction des classificateurs pour la traduction humaine (HUM), la seconde (**Label 1**), les prédictions pour la traduction neuronale (NMT). Vous pourriez utiliser cette information (typiquement le nombre de classificateurs faisant la bonne prédiction, ce qui est indiqué après *tot* :) pour vous guider dans l'analyse des traductions (l'exemple montre un cas où les classificateurs identifient clairement une signature de traduction neuronale).

SRC	Canadians have come to the full realization that we are too small a nation to simply to trade with each other.
HUM	Les Canadiens se rendent bien compte que notre pays est trop petit pour que les provinces se livrent au commerce uniquement les unes avec les autres.
NMT	Les Canadiens ont pris pleinement conscience que nous sommes une trop petite nation pour simplement commercer entre eux.
Label 0	000010001000001000 mono : 7 bi : 8 tot : 15
Label 1	101111101111111110 mono : 7 bi : 8 tot : 15

Vous devez produire un **rapport** d'au plus 3 pages (format pdf, en français ou en anglais) résumant vos découvertes.

Remise

La remise est à faire sur Studium sous le libellé **devoir9**. Vous devez remettre votre rapport (format pdf, texte en anglais ou en français) ainsi que les ressources annotées demandées dans une archive (gzip, tar, tar.gz) dont le nom est préfixé de **devoir9-name1** ou **devoir9-name1-name2** selon que vous remettez seul ou à deux, où **name1** et **name2** sont à remplacer par l'identité des personnes faisant la remise (**prénom_nom**). Assurez vous que le nom des personnes impliquées dans le devoir soit indiqué sur tous les documents remis (code et ressources). Le devoir est à remettre en groupe d'au plus deux personnes au plus tard vendredi 4 décembre à 23h59.