

TP individuel sur la traduction automatique

Dans le cadre du module *Problèmes, Méthodes et Outils pour le TALN*

À rendre pour le 7 janvier 2022 à 12 heures.

Objectif

Vous sensibiliser à la traduction humaine, à la traduction automatique et à son évaluation.

Données

Vous devez choisir un texte de 250 mots rédigé dans une langue que vous connaissez bien et qui n'est pas votre langue maternelle¹. Ce texte doit traiter d'un sujet que vous connaissez bien. Ce peut être un article de journal ou de revue disponible sous forme électronique.

Vous traduirez ce texte dans votre langue maternelle en deux temps : une traduction que vous ferez vous-même, une traduction que vous produirez en post-éditant des résultats de traduction automatique.

Traduction humaine

Vous traduirez par vous-même la première moitié du texte (≈125 mots) avec comme seul aide autorisée :

- Un dictionnaire bilingue (je vous recommande WordReference.com²),
- Un concordancier ou une mémoire de traduction (je vous recommande Linear B³ et/ou Lingee⁴)

Vous rangerez cette moitié du texte dans un tableau comme suit :

Phrase source	Traduction	Temps passé	Remarques
This is the first sentence.	Ceci est la première phrase.	5 secondes	Aucune difficulté
The inner housing is located between the engine valve and camshaft and transmits the zero or partial lift to the valve.	Le boîtier intérieur est situé entre la soupape du moteur et l'arbre à cames et transmet à la soupape la levée nulle ou partielle.	3 minutes	Recherche de "housing", "camshaft" et "valve" dans le dictionnaire. En particulier pour un moteur ("engine") le terme "valve" se traduit par « soupape »

Tableau 1 : Organisation de la présentation du travail de traduction manuelle

Une ligne par phrase avec sa traduction, le temps passé et des remarques sur votre travail.

Vous commencerez la traduction quand la première colonne du tableau sera prête. Vous calculerez le temps total passé à la traduction de ces 125 mots et donnerez la moyenne de temps par mot.

Quand la traduction de chaque phrase sera terminée, vous copierez toutes les traductions les unes à la suite des autres pour faire une première version de traduction. Vous copierez cette première version

¹ Si vous n'êtes pas francophone, je préfère que vous choisissiez un texte dans votre langue maternelle et que vous fassiez le travail demandé vers le français ou l'anglais (la langue que vous maîtrisez le mieux).

² <http://www.wordreference.com>

³ <http://www.linearb.co.uk/>

⁴ <http://www.linguee.fr/>

de traduction pour réviser le texte globalement produit en mettant en évidence ce que vous avez changé par rapport à la première traduction.

Première Traduction

Bla bla bla. Ble ble bla.

Traduction révisée

Bla bla bla. Ble **bla** bla.

Traduction automatique

Vous allez obtenir différents résultats de traduction automatique pour la seconde moitié du texte (≈125 mots), je vous conseille d'utiliser au moins :

- Google translate⁵ (de nombreux couples de langues)
- SYSTRANet⁶ (de nombreux couples de langues)

et aussi d'autres sites que je vous propose⁷.

Vous rangerez cette moitié du texte dans un tableau comme suit :

	Propositions			Ma post-édition			
Phrase source	Google	Systran	Prompt	Choix	Post-édition	Temps passé	Remarques
This is the first sentence.	Il s'agit de la première phrase.	C'est la première phrase.	C'est la première phrase.	Systran	C'est la première phrase.	0 seconde	Aucune difficulté
The inner housing is located between the engine valve and camshaft and transmits the zero or partial lift to the valve.	Le boîtier intérieur est situé entre la soupape de moteur et l'arbre à cames et transmet la levée nulle ou partielle de la vanne .	Le logement intérieur est placé entre la valve de réacteurs et l'arbre à cames et transmet l'ascenseur zéro ou partiel à la valve.	Le logement intérieur est localisé entre la valve de locomotive et l'arbre à cames et transmet l'ascenseur zéro ou partiel à la valve.	Google	Le boîtier intérieur est situé entre la soupape de moteur et l'arbre à cames et transmet la levée nulle ou partielle à la soupape.	30 secondes	Les choix de vocabulaire pour Systran et Prompt sont mauvais : logement, réacteur, ascenseur. « de la vanne » transformé en « à la soupape ». Vérification du choix de « soupape » dans le dico

Tableau 2 : Organisation de la présentation du travail de post-édition

Une ligne par phrase avec les propositions de traduction de différents systèmes (au moins 2), le choix de la meilleure traduction, sa post-édition (une traduction correcte en changeant au minimum la traduction choisie), le temps passé et des remarques sur votre travail (raison de votre choix de traduction à post-éditer, description des modifications faites).

Vous commencerez la post-édition quand vous aurez rempli, pour chaque phrase, les propositions de traduction. Vous calculerez le temps total passé à la traduction de ces 125 mots et donnerez la moyenne de temps par mot. Vous comparerez ce chiffre à celui que vous avez trouvé pour la première partie de l'expérience. Que pouvez-vous en conclure ?

Quand la traduction de chaque phrase sera terminée, vous copierez toutes les traductions les unes à la suite des autres pour faire une première version de traduction. Vous copierez cette première version de traduction pour réviser le texte complet produit en mettant en évidence ce que vous avez changé par rapport à la traduction post-éditée.

⁵ <http://translate.google.com>

⁶ <http://www.systranet.com/>

⁷ Voir la dernière section de ce document

Première Traduction

Bla bla bla. Ble ble bla.

Traduction révisée

Bla bla bla. Ble **bla** bla.

Évaluation automatique

Il vous faut aussi calculer le Translation Error Rate (TER) pour chacune des post-éditions que vous avez produites. Pour ce faire il vous faut utiliser le logiciel de calcul qui est fourni dans la rubrique « TD & TP » du site Web de l'enseignement.

Vous téléchargerez d'abord le fichier « ter-0.7.25.zip ». Après décompression vous obtiendrez un dossier qui contient le programme java exécutable et un fichier « README.txt » qui contient toutes les instructions d'utilisation.

Vous construirez 2 fichiers alignés (Eval_Ter_Hypos.txt & Eval_Ter_Refs.txt) avec respectivement sur chaque ligne la **proposition** de traduction que vous avez post-éditer et la **post-édition** que vous proposez. Vous pourrez alors calculer le score TER pour chaque phrase.

Vous reporterez ces résultats dans un tableau comme suit :

Phrase source	Proposition ⁸	Post-édition ⁹	Score TER
Phrase 1	Traduction 1	Post-édition 1	TER1 %
Phrase 2	Traduction 2	Post-édition2	TER2 %

Tableau 3 : Présentation des résultats du calcul du TER

Qu'est-ce que ces chiffres vous apprennent sur les traductions que vous avez post-édité ?

Conclusion

Dans une brève conclusion vous direz ce que vous avez appris au travers de cette expérience.

Rendu

Vous m'enverrez à mon adresse mail (Herve.Blanchon@univ-grenoble-alpes.fr) un dossier compressé (zip) contenant le texte de votre rapport, en format pdf, ainsi que les fichiers que vous avez produits pour le calcul du TER et le fichier de résultats. Le fichier compressé s'intitulera **WTrad_3 premières lettres de votre nom_3 premières lettres de votre prénom** (pour moi : **WTrad_Bla_Her.zip**).

Le titre de votre message sera

[PMOTALN] TP Trad – votre_nom votre_prénom

(pour moi : **[PMOTALN] TP TRAD – Blanchon Hervé**).

N'oubliez pas de mettre votre nom dans votre rapport et de lui donner un titre.

Date de rendu : le 7 janvier 2022 à 12 heures.

Site web du module

<http://lig-membres.imag.fr/blanchon/SitesEns/PMOTALN/>

⁸ Traduction que vous aurez choisie dans la colonne propositions du Tableau 2.

⁹ Post-édition que vous aurez proposée dans la colonne post-édition du Tableau 2.

Systèmes disponibles en ligne¹⁰

Provider	URL	Source Languages	Target Languages
Google	https://translate.google.com/	105 languages	105 languages
DeepL	https://www.deepl.com/translator	French, English, German, Spanish, Italian, Dutch, Polish	
Systran	http://www.systranet.com/translate	Arabic, Chinese, Dutch, English, French, German, Greek, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish, Swedish	Source language dependent
PROMT ¹¹	http://www.online-translator.com	Arabic, Catalan, Chinese, English, French, Finnish, German, Greek, Italian, Japanese, Kazakh, Korean, Polish, Portuguese, Russian, Spanish, Turkish	Source language dependent
Lingue.tech*	https://www.lingue.tech/de/en/personal-translator-demo/	Arabic, Chinese, English, French, German, Italian, Portuguese, Spanish	Source language dependent
Morphologic*	http://www.webforditas.hu/?show=textTab&lang=english	Bulgarian, Danish, English, French, German, Hungarian, Latvian, Norwegian, Polish, Portuguese, Russian, Spanish, Ukrainian	
SkyCode	http://webtrance.skycode.com/online.asp?setlang=en	Bulgarian, English, French, German, Italian, Spanish	Bulgarian, Chinese, English, French, German, Italian, Spanish, Turkish
Amebis	http://presis.amebis.si/prevajanje/?jezik=en&	German ⇒ Slovenian, English ⇒ Slovenian, Slovenian ⇒ English	
PWN	http://translatca.pl	English, German, Polish, Russian	
Lingenio	http://wwwwtranslate.eu	English ⇔ German, English ⇔ French, English ⇔ Chinese	

¹⁰ Tous les sites web de traduction gratuits ne sont pas répertoriés ici, vous pouvez également chercher par vous-même

¹¹ L'utilisateur peut sélectionner un domaine pour le texte source