Classification automatique des procédés de traduction

Yuming Zhai, Gabriel Illouz, Anne Vilnat

Classification automatique des procédés de traduction

Yuming Zhai Gabriel Illouz Anne Vilnat
LIMSI-CNRS, Univ. Paris-Sud, Univ. Paris-Saclay, France
{prénom.nom}@limsi.fr

RÉSUMÉ

En vue de distinguer la traduction littérale des autres procédés de traduction, des traducteurs et linguistes ont proposé plusieurs typologies pour caractériser les différents procédés de traduction, tels que l'équivalence idiomatique, la généralisation, la particularisation, la modulation sémantique, etc. En revanche, les techniques d'extraction de paraphrases à partir de corpus parallèles bilingues n'ont pas exploité ces informations. Dans ce travail, nous proposons une classification automatique des procédés de traduction en nous basant sur des exemples annotés manuellement dans un corpus parallèle (anglais-français) de *TED Talks*. Même si le jeu de données est petit, les résultats expérimentaux sont encourageants, et les expériences montrent la direction à suivre dans les futurs travaux.

ABSTRACT

Automatic Classification of Translation Processes

In order to distinguish literal translation from other translation processes, translators and linguists have proposed several typologies to characterize different translation processes, such as idiomatic equivalence, generalization, particularization, semantic modulation, etc. However, the techniques to extract paraphrases from bilingual parallel corpora have not exploited this information. In this work, we propose an automatic classification of translation processes, based on manually annotated examples in an English-French parallel corpus of *TED Talks*. Even with a small dataset, the experimental results are encouraging and our experiments show the direction to follow in future work.

MOTS-CLÉS : procédés de traduction, classification automatique, extraction de paraphrases.

KEYWORDS: translation processes, automatic classification, paraphrase extraction.

1 Introduction

Les procédés de traduction sont étudiés depuis longtemps (Vinay & Darbelnet, 1958; Chuquet & Paillard, 1989; Molina & Hurtado Albir, 2002). Ils distinguent les traductions littérales des autres procédés de traduction au niveau sous-phrastique. Prenons comme exemple ces deux traductions humaines non littérales : la première traduction préserve exactement le sens, où l'expression figée *à la hauteur de* possède un sens figuré *« capable de résoudre »* ; en revanche, la deuxième traduction est plus compliquée, où il existe une inférence textuelle entre le segment source et la traduction.

(1.EN) a solution ***that's big enough to solve*** *our problems*
(1.FR) une solution ***à la hauteur de*** *nos problèmes*
(2.EN) and ***that scar has stayed with him*** *for his entire life*
(2.FR) et que, toute sa vie, ***il a souffert de ce traumatisme***

Les traductions non littérales peuvent poser des difficultés pour l'alignement de mots automatique (Dorr *et al.*, 2002; Deng & Xue, 2017), ou causer des changements de sens dans certains cas. Cependant, à notre connaissance, les techniques de traitement automatique des langues n'ont pas explicitement exploité ces procédés de traduction. Bannard & Callison-Burch (2005) ont proposé d'exploiter les techniques de traduction automatique pour extraire des paraphrases à partir de corpus parallèles bilingues. Leur hypothèse est que deux segments monolingues sont des paraphrases potentielles s'ils partagent des traductions communes dans une autre langue. Actuellement, la plus grande ressource de paraphrases, PPDB (ParaPhrase DataBase) (Ganitkevitch *et al.*, 2013; Pavlick *et al.*, 2015b), a été construite selon cette méthode. En revanche, Pavlick *et al.* (2015a) ont révélé qu'il existe d'autres relations que l'équivalence stricte (paraphrase) dans PPDB (*i.e. Implication (dans les deux directions), Exclusion, Lié à* et *Indépendant*) [1]. Des traductions « pivots » non littérales dans des corpus parallèles bilingues peuvent influencer l'équivalence stricte entre les candidats de paraphrases extraits, néanmoins elles n'ont pas reçu assez d'attention pendant cette exploration de corpus.

De notre côté, en nous basant sur les théories développées en traduction, nous avons annoté et analysé manuellement un corpus parallèle anglais-français de *TED Talks*. Ce travail nous permet de proposer une typologie de procédés de traduction adaptée à notre corpus, ainsi que d'établir le guide d'annotation. Dans cet article, nous présentons une classification automatique des procédés de traduction en utilisant ce corpus annoté. Après avoir présenté les travaux précédents (section 2), nous décrivons l'annotation manuelle et le jeu de données (section 3). La section 4 présente les traits exploités pour la classification automatique et la section 5 montre les résultats et les analyse. La conclusion et les perspectives suivent dans la section 6.

2 Travaux précédents

Vinay & Darbelnet (1958) ont identifié des procédés de traduction directe et indirecte, ces derniers correspondent aux cas où une traduction littérale est inacceptable, ou lorsque les asymétries structurelles ou conceptuelles entre la langue source et la langue cible ne sont pas négligeables. Ces travaux ont été poursuivis par Newmark (1981, 1988) et Chuquet & Paillard (1989). Plus récemment, Molina & Hurtado Albir (2002) ont proposé leur propre classification basée sur une étude de la traduction des éléments culturels du roman *Cent ans de solitude* de l'espagnol vers l'arabe. Pour le couple anglais-chinois, Deng & Xue (2017) identifient, catégorisent et quantifient semi-automatiquement sept types de divergences de traduction, causées par des traductions non littérales ou des différences grammaticales inter-linguistiques [2]. Nous annotons le corpus selon une typologie inspirée par ces travaux précédents, mais aussi adaptée au corpus de *TED Talks*.

Récemment, différents modèles ont été proposés pour détecter automatiquement des divergences en traduction dans des corpus parallèles. Le but est de filtrer automatiquement des couples de phrases divergents afin d'améliorer la performance des systèmes de traduction automatique. Carpuat *et al.* (2017) ont introduit un détecteur de divergence cross-lingue basé sur SVM, en utilisant des traits en alignement de mots et en longueur de phrase. Vyas *et al.* (2018) ont proposé une approche basée sur des réseaux neuronaux profonds, et l'entraînement ne demande pas d'annotation manuelle. D'une façon non supervisée, Pham *et al.* (2018) ont généré des plongements phrastiques en fonction de

1. Exclusion : X est le contraire de Y ; X et Y s'excluent mutuellement. Lié à : X est lié d'une certaine manière à Y. (*p. ex. pays / patriotique*). Indépendant : X n'est pas lié à Y.

2. Encodage lexical ; différence de transitivité ; absence de mots grammaticaux spécifiques à une langue ; différence de catégories de phrases ; différence dans l'ordre de mots ; éléments omis ; paraphrases structurelles.

la similarité entre les mots. Ils mesurent l'équivalence sémantique entre les phrases afin de guider le filtrage. Contrairement à ces efforts qui ont lieu au niveau phrastique et effectuent une décision binaire, nous classifions automatiquement différents procédés de traduction au niveau sous-phrastique à partir d'exemples annotés manuellement. Ceci permettra d'identifier certains procédés de traduction qui peuvent provoquer des divergences sémantiques, tandis que d'autres conservent le sens original.

3 Annotation manuelle et description des données

Afin de modéliser les choix de traduction effectués par les traducteurs humains au niveau sous-phrastique, nous avons annoté un corpus parallèle trilingue (anglais-français, anglais-chinois) de *TED Talks* [3] en procédés de traduction (Zhai, 2018; Zhai *et al.*, 2018). Le corpus est composé de transcriptions et de traductions humaines de présentations orales. Nous choisissons ce genre spécifique parce que plusieurs domaines sont couverts dans les présentations, et la diversité des phénomènes de traduction est entre celle présente dans les corpus littéraires et techniques. L'accord inter-annotateur Kappa (Cohen, 1960) pour annoter le corpus de contrôle anglais-français et anglais-chinois est de 0,67 et 0,61, tous proches du seuil pour être suffisant [4]. Cela indique que la tâche de l'annotation manuelle est complexe.

Nous présentons dans la table 1 une brève définition, un exemple typique et le nombre d'instances pour chaque catégorie à classifier automatiquement pour le couple anglais-français [5]. Nous combinons *transposition* et *mod+trans* dans une catégorie *contient_transposition*, où la classe *modulation* est considérée comme neutre. Dans ce présent travail, nous menons des expériences dans un scénario simplifié, où nous connaissons déjà les frontières des couples bilingues, et nous ne prédisons que le procédé de traduction. Par exemple, étant donné le couple *deceptive* → *une illusion*, le but est de prédire son étiquette *contient_transposition*.

4 L'ingénierie des traits pour la classification automatique

Nous avons exploité quatre groupes de traits ci-dessous pour le couple anglais-français. Les jeux d'étiquettes des deux langues pour l'analyse morpho-syntaxique, l'analyse syntaxique en constituant et en dépendance ont été convertis en trois jeux unifiés et compacts (Petrov *et al.*, 2012).

Analyse morpho-syntaxique (PoS) 1) L'analyse est faite par *Stanford CoreNLP* (Manning *et al.*, 2014) pour les deux langues. Pour chaque langue, le nombre d'occurrence de chaque étiquette est compté dans un vecteur. Nous calculons aussi la similarité cosinus entre ces deux vecteurs (sur tous les mots et sur seulement les mots pleins [6]).
2) Nous vérifions le patron de changement de séquence de PoS selon une liste construite manuellement. Par exemple le couple *methodologically* → *de façon méthodologique* correspond au patron *ADV* → *ADP NOUN ADJ*.

3. `https://www.ted.com/`
4. La valeur minimum pour atteindre un accord suffisant est de 0,61.
5. Notez qu'il existe d'autres règles d'annotation détaillées dans le guide d'annotation.
6. Les étiquettes de mots pleins contiennent : ADJ, ADV, NOUN, PROPN, VERB. Si un segment ne contient aucun mot plein, nous utilisons le segment original.

Procédé	Définition et exemple typique
littéral (3771)	Traduction mot à mot. *certain **kinds** of → certains **types** de*
équivalence (289)	Traduction non littérale des proverbes ou des expressions figées ; Une traduction mot à mot est possible mais le traducteur exprime différemment, sans changer le sens ni les catégories grammaticales. ***back then → à l'époque***
transposition (289)	Modification des catégories grammaticales sans en changer le sens. *unless **something changes** → à moins qu'**un changement ait lieu***
modulation (195)	Modulation métonymique et grammaticale (Chuquet & Paillard, 1989) ; Changement du point de vue ; Changement du sens possible. ***that scar has stayed with him → il a souffert de ce traumatisme***
mod+trans (53)	Combinaison des transformations de *Modulation* et de *Transposition*, ce qui peut rendre l'alignement de mots difficile. *this is a completely **unsustainable** pattern → il est absolument **impossible de continuer sur** cette tendance*
généralisation (86)	Plusieurs mots ou expressions sources peuvent être traduits en un mot ou une expression cible avec un sens plus général, le traducteur utilise ce dernier. *as we **sit here** in ... → alors que nous **sommes** à ...*
particularisation (215)	Le mot ou l'expression source peut être traduit en plusieurs mots ou expressions cibles avec un sens plus spécifique. Le traducteur en choisit un selon le contexte. *they **have** a screen → ils **sont équipés d'**un écran*

TABLE 1 – Définition, exemple typique et nombre d'instances des procédés de traduction à classifier automatiquement.

Surface 3) Le nombre de tokens dans les deux segments (l_e, l_f), le ratio de ces nombres (l_e/l_f, l_f/l_e), la distance Levenshtein (Levenshtein, 1966) entre les segments.

Analyse syntaxique 4) L'analyse syntaxique en constituant est faite par *Bonsai* (Candito *et al.*, 2010) pour le français, par *Stanford CoreNLP* pour l'anglais [7]. Nous comparons les étiquettes PoS pour un couple de mots ; les étiquettes du nœud non terminal pour un couple de segments ; la catégorie des étiquettes (*i.e.* verbe → syntagme verbal) pour un mot traduit par un segment ou vice versa.
5) L'analyse syntaxique en dépendance est faite par *Stanford CoreNLP* pour les deux langues afin de partager le même jeu d'étiquettes. À l'intérieur des segments, nous comptons le nombre d'occurrence de chaque relation de dépendance. À l'extérieur des segments, parmi les mots liés en dépendance dans chaque langue, nous gardons ceux qui sont manuellement alignés préalablement. Ensuite, nous comptons le nombre d'occurrence de chaque relation de dépendance que les mots à l'intérieur du segment entretiennent avec ces mots de contexte.

Ressource externe 6) Nous calculons la similarité cosinus entre les plongements provenant de *ConceptNet Numberbatch* (Speer *et al.*, 2017). Cette ressource est multilingue et le système basé sur *ConceptNet* a remporté la première place dans la tâche "Similarité sémantique lexicale multilingue et cross-lingue" de SemEval2017 (Camacho-Collados *et al.*, 2017; Speer & Lowry-Duda, 2017).

7. Pour l'analyse en constituant, *Bonsai* est beaucoup plus rapide que *Stanford CoreNLP* et a moins d'erreurs évidentes.

Certaines expressions multi-mots ont leur propre plongement dans cette ressource. Sinon, nous calculons la moyenne des plongements sur seulement les mots pleins. Les mêmes traits ont été calculés pour les segments lemmatisés [8].
7) La ressource *ConceptNet* fournit des assertions sous forme de triplet : un couple de mots ou expressions liés par une relation [9]. Dans cette ressource multilingue, nous vérifions si un couple anglais-français est directement lié ; indirectement lié par un autre segment français ou simplement pas lié. Trois formes sont testées : forme originale, forme lemmatisée et forme lemmatisée filtrée. [10]
8) Sur la forme lemmatisée filtrée, nous calculons le pourcentage des tokens bilingues qui sont liés avec une relation de dérivation, en basant sur la ressource *ConceptNet*. Par exemple *deceptive* et *illusion* ne sont pas directement liés dans la ressource, mais tous les deux sont liés à *illusoire*. Ainsi nous considérons qu'il existe un lien de dérivation entre eux.

Alignement de mot Pour ce groupe de traits, nous avons exploité la table de probabilité de traduction lexicale générée par l'outil statistique d'alignement de mots *Berkeley Word Aligner* (Liang *et al.*, 2006), entraîné sur un corpus parallèle anglais-français combiné de *TED Talks* et d'une partie du corpus Paracrawl [11] (au total 1.8M de couples de phrases et 41M de tokens anglais) :

9) L'entropie des distributions de probabilités de traduction lexicale (Gray, 1990; Carl & Schaeffer, 2017) : calculée selon cette équation : $H(X) = \sum_i P(x_i)I(x_i) = -\sum_i P(x_i)log_e P(x_i)$. Nous calculons l'entropie moyenne sur des mots pleins. Une entropie plus grande indique que les mots possèdent des sens plus généraux ou qu'ils sont polysémiques. Le même trait est calculé sur les mots pleins lemmatisés.
10) La pondération lexicale bidirectionnelle sur les mots pleins, en supposant un alignement de mots *n-m* (a) entre deux segments ($\bar{e}$ et $\bar{f}$). Selon l'équation proposée par Koehn *et al.* (2003) [12], pour calculer la pondération lexicale directe, chacun des mots anglais e_i est généré par des mots étrangers alignés f_j avec la probabilité de traduction lexicale $w(e_i|f_j)$. Et de même pour la pondération lexicale inverse $lex(\bar{f}|\bar{e}, a)$. Les mêmes traits ont été calculés pour les mots pleins lemmatisés. Ce trait pourrait refléter la confiance de l'alignement entre les deux segments.
11) La somme de différence de probabilités de traduction lexicale entre la traduction humaine et la traduction la plus probable selon la table de probabilité. Pour chaque mot source, nous prenons le mot cible dans la traduction humaine avec la plus grande probabilité. Par exemple pour la paire *alternatives* → *solutions de remplacement*, la traduction la plus littérale est *alternatives* avec une probabilité de 0,4. Dans la traduction humaine, le mot *solutions* possède la plus grande probabilité, mais qui est seulement 0,07. Selon cette méthode, nous comptons aussi les mots non alignés pour calculer un ratio sur le nombre total de tokens de chaque côté. Ces traits ont été calculés dans les deux directions de traduction.

Le nombre d'instances pour la validation croisée est assez limité, nos expériences utilisant des classifieurs d'apprentissage statistique obtiennent de meilleurs résultats qu'avec les réseaux neuronaux (Zhai *et al.*, 2019). La boîte à outils *Scikit-Learn* (Pedregosa *et al.*, 2011) est utilisée pour entraîner différents classifieurs [13].

8. La lemmatisation anglaise est faite par *Stanford CoreNLP*, celle française par *Tree Tagger* (Schmid, 1995), puisque ce n'est pas encore possible par *Stanford CoreNLP*.
9. `https://github.com/commonsense/conceptnet5/wiki/Downloads`
10. Nous filtrons les mots selon une liste manuelle, qui contient les verbes légers, déterminants, pronoms, etc.
11. `https://wit3.fbk.eu/`, `https://paracrawl.eu/index.html`
12. $lex(\bar{e}|\bar{f}, a) = \prod_{i=1}^{length(\bar{e})} \frac{1}{|\{j|(i,j)\in a\}|} \sum_{\forall(i,j)\in a} w(e_i|f_j)$
13. Le jeu de données et le code sont disponibles ici : `https://github.com/YumingZHAI/ctp`.

5 Résultats expérimentaux et analyse

Le nombre d'instances de *non_littéral* (1127) est seulement un tiers de *littéral* (3771). Compte tenu de cet écart, nous avons évalué les classifieurs sous plusieurs configurations : (*a*) six classes (*littéral, équivalence, généralisation, particularisation, modulation, contient_transposition*), où *littéral* contient toutes les instances, ou 200 instances pour une distribution approximativement équilibrée [14]. (*b*) deux classes (*littéral* et *non_littéral*), avec trois répartitions (3 :1, 2 :1, 1 :1) (*c*) cinq classes : seulement les catégories non littérales.

Ces classifieurs ont été entraînés : *RandomForest, Multilayer Perceptron, Logistic Regression, Support Vector Machine, K-nearest Neighbors, Decision Tree, Bernoulli Naive Bayes, multinomial Naive Bayes* et *Gaussian Naive Bayes*. Pour chaque configuration, nous avons optimisé les hyperparamètres de ces classifieurs [15]. L'évaluation est menée par une validation croisée à cinq plis (en utilisant *StratifiedKFold*), selon les mesures de l'exactitude (*accuracy*) moyenne, la F-mesure micro-moyenne et macro-moyenne (Tsoumakas *et al.*, 2011). Les résultats sous différentes configurations sont récapitulés dans la table 2, où le classifieur *Dummy* est utilisé comme une baseline, qui prédit toujours la classe la plus nombreuse. Pour toutes les configurations, le classifieur *RandomForest* obtient toujours la meilleure performance [16].

Nous essayons d'abord une classification directe en six classes. Les résultats de notre classifieur dépassent largement ceux du classifieur *Dummy*. En revanche, la difficulté de la tâche en multi-classe est aussi reflétée dans la distribution approximativement équilibrée. Ainsi nous décidons de diviser le problème : effectuer d'abord une classification binaire, suivi par une classification multi-classe parmi les catégories non littérales.

Pour la classification binaire, les deux meilleurs classifieurs sont *RandomForest* et *Multilayer Perceptron*. En plus, *RandomForest* est meilleur que les deux assemblés par la méthode *hard voting* ou *soft voting*. De la distribution naturelle (3 :1) à la distribution équilibrée (1 :1), la F-mesure moyenne pour la classe *non_littéral* augmente de 0,78 à 0,88. Nous continuerons à tester cette tendance quand un jeu de données plus large sera disponible. Une analyse des erreurs sur la distribution 3 :1 montre que parmi les 290 instances *non_littéral* classifiées en *littéral*, 117 sont de classe *équivalence*. Cela indique que ces deux classes sont difficiles à distinguer pour le classifieur.

L'exactitude la plus élevée pour la classification entre les classes non littérales est de 55,10%. Des F-mesure moyennes sur les cinq plis pour chaque classe sont : *équivalence* (0,51), *généralisation* (0,25), *particularisation* (0,56), *modulation* (0,36) et *contient_transposition* (0,68). La catégorie *généralisation* contient beaucoup moins d'instances que les autres catégories, qui nécessite une augmentation ; il existe beaucoup de confusion entre *modulation* et les autres catégories, qui suggère une amélioration du guide d'annotation ; la confusion existe aussi entre *équivalence* et *contient_transposition*.

Avec le meilleur classifieur *RandomForest*, nous avons effectué une étude d'ablation de traits. Pour la classification binaire, le groupe de traits *alignement de mot* contribue le plus. Pour la classification en cinq classes, la combinaison de tous les traits sauf le groupe *ressource externe* génère le meilleur résultat (exactitude moyenne 55,20%), où les groupes *analyse morpho-syntaxique* et *analyse syntaxique* contribuent plus. En général, des traits en nombre réel ont des meilleures performances que

14. Des instances de *littéral* ont été extraites au hasard pour les configurations *a* et *b*.

15. Pour trouver les meilleurs hyperparamètres, 10% de données sont séparées comme test, et une validation croisée à trois plis est exécutée sur 90% de données d'entraînement.

16. Les hyperparamètres en détail sont donnés ensemble avec le code.

des traits en nombre entier.

Distribution de classes	Classifieur	Exactitude moyenne	Micro-F1	Macro-F1
Six classes				
six classes, avec 3771 *littéral*	Dummy	76,99%	0,77	0,14
	RandomForest	**83,10%**	0,83	0,44
six classes, avec 200 *littéral*	Dummy	25,77%	0,26	0,07
	RandomForest	57,04%	0,57	0,52
Deux classes				
littéral (3) : *non_littéral* (1)	Dummy	76,99%	0,77	0,43
	RandomForest	**90,16%**	0,90	0,86
littéral (2) : *non_littéral* (1)	Dummy	66,67%	0,67	0,40
	RandomForest	88,85%	0,89	0,88
littéral (1) : *non_littéral* (1)	Dummy	50,00%	0,50	0,33
	RandomForest	**87,09%**	0,87	0,87
Cinq classes				
Cinq classes non_littéral	Dummy	30,35%	0,30	0,09
	RandomForest	**55,10%**	0,55	0,47

TABLE 2 – Résultats expérimentaux sous différentes configurations, utilisant tous les traits

6 Conclusion et perspectives

En nous fondant sur notre corpus annoté manuellement en procédés de traduction au niveau sous-phrastique, nous avons proposé une classification automatique. Avec les traits implémentés et par le classifieur *RandomForest*, l'exactitude la plus élevée est de 87,09% pour la classification binaire (distribution équilibrée), et de 55,20% pour la classification entre les procédés de traduction non littérale. Les résultats de notre classifieur sont encourageants et nous continuerons à l'améliorer. Nous utiliserons cette connaissance pour mieux contrôler le processus d'extraction de paraphrases à partir de corpus parallèle bilingues. Il est aussi pertinent de l'intégrer dans le processus de traduction automatique pour mieux traiter les traductions non littérales, ou de l'utiliser pour assister aux études en traductologie.

La classe *Généralisation* contient beaucoup moins d'instances que les autres classes. Nous aurons recours à la ressource PPDB, ConceptNet et Linguee pour construire un jeu de données plus équilibré. L'annotation manuelle se poursuivra pour fournir plus de données, surtout pour le couple anglais-chinois. Nous exploiterons d'autres traits pour mieux effectuer la classification multi-classe, tels que la probabilité de traduction des segments, la liste des expressions figées, etc. Une perspective importante est d'étendre ce travail sur des corpus parallèles non alignés manuellement au préalable. Cette configuration demandera un alignement de mot automatique de bonne performance : d'abord aligner les traductions littérales mot à mot, ensuite aligner des blocs *n-m* sur des couples de segments traduits non littéralement.

Remerciements

Nous remercions les relecteurs anonymes pour leurs nombreuses remarques constructives. Nous exprimons aussi notre gratitude à Aurélien Max pour ses propositions des traits pertinents à implémenter, et à Cyril Grouin pour ses conseils en vue d'améliorer le guide d'annotation.

Références

BANNARD C. & CALLISON-BURCH C. (2005). Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, p. 597–604 : Association for Computational Linguistics.

CAMACHO-COLLADOS J., PILEHVAR M. T., COLLIER N. & NAVIGLI R. (2017). Semeval-2017 task 2 : Multilingual and cross-lingual semantic word similarity. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, p. 15–26 : Association for Computational Linguistics.

CANDITO M., NIVRE J., DENIS P. & ANGUIANO E. H. (2010). Benchmarking of statistical dependency parsers for French. In *Proceedings of the 23rd International Conference on Computational Linguistics : Posters*, p. 108–116 : Association for Computational Linguistics Chinese Information Processing Society of China.

CARL M. & SCHAEFFER M. J. (2017). Why Translation Is Difficult : A Corpus-Based Study of Non-Literality in Post-Editing and From-Scratch Translation. *HERMES-Journal of Language and Communication in Business*, (56), 43–57.

CARPUAT M., VYAS Y. & NIU X. (2017). Detecting cross-lingual semantic divergence for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, p. 69–79 : Association for Computational Linguistics.

CHUQUET H. & PAILLARD M. (1989). *Approche linguistique des problèmes de traduction anglais-français*. Ophrys.

COHEN J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

DENG D. & XUE N. (2017). Translation Divergences in Chinese-English Machine Translation : An Empirical Investigation. *Computational Linguistics*, **43**(3), 521–565.

DORR B. J., PEARL L., HWA R. & HABASH N. (2002). Duster : A method for unraveling cross-language divergences for statistical word-level alignment. In *Conference of the Association for Machine Translation in the Americas*, p. 31–43 : Springer.

GANITKEVITCH J., VAN DURME B. & CALLISON-BURCH C. (2013). PPDB : The paraphrase database. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, p. 758–764.

GRAY R. M. (1990). *Entropy and Information Theory*. Berlin, Heidelberg : Springer-Verlag.

KOEHN P., OCH F. J. & MARCU D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, p. 48–54 : Association for Computational Linguistics.

LEVENSHTEIN V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, **10**(8), 707–710.

LIANG P., TASKAR B. & KLEIN D. (2006). Alignment by agreement. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, p. 104–111 : Association for Computational Linguistics.

MANNING C. D., SURDEANU M., BAUER J., FINKEL J., BETHARD S. J. & MCCLOSKY D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, p. 55–60.

MOLINA L. & HURTADO ALBIR A. (2002). Translation Techniques Revisited : A Dynamic and Functionalist Approach. *Meta*, **47**(4), 498–512.

NEWMARK P. (1981). *Approaches to Translation (Language Teaching Methodology Senes)*. Oxford : Pergamon Press.

NEWMARK P. (1988). *A textbook of translation*, volume 66. Prentice Hall New York.

PAVLICK E., BOS J., NISSIM M., BELLER C., VAN DURME B. & CALLISON-BURCH C. (2015a). Adding semantics to data-driven paraphrasing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1 : Long Papers)*, volume 1, p. 1512–1522.

PAVLICK E., RASTOGI P., GANITKEVITCH J., DURME B. V. & CALLISON-BURCH C. (2015b). PPDB 2.0 : Better Paraphrase Ranking, Fine-Grained Entailment Relations, Word Embeddings, and Style Classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 2 : Short Papers*, p. 425–430.

PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTENHOFER P., WEISS R., DUBOURG V., VANDERPLAS J., PASSOS A., COURNAPEAU D., BRUCHER M., PERROT M. & DUCHESNAY E. (2011). Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

PETROV S., DAS D. & MCDONALD R. T. (2012). A Universal Part-of-Speech Tagset. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, M. U. DOGAN, B. MAEGAARD, J. MARIANI, J. ODIJK & S. PIPERIDIS, Eds., *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*, p. 2089–2096 : European Language Resources Association (ELRA).

PHAM M. Q., CREGO J., SENELLART J. & YVON F. (2018). Fixing Translation Divergences in Parallel Corpora for Neural MT. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, p. 2967–2973.

SCHMID H. (1995). Improvements In Part-of-Speech Tagging With an Application To German. In *Proceedings of the ACL SIGDAT-Workshop*, p. 47–50.

SPEER R., CHIN J. & HAVASI C. (2017). Conceptnet 5.5 : An open multilingual graph of general knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*, p. 4444–4451.

SPEER R. & LOWRY-DUDA J. (2017). Conceptnet at semeval-2017 task 2 : Extending word embeddings with multilingual relational knowledge. In S. BETHARD, M. CARPUAT, M. APIDIANAKI, S. M. MOHAMMAD, D. M. CER & D. JURGENS, Eds., *Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, Canada, August 3-4, 2017*, p. 85–89 : Association for Computational Linguistics.

TSOUMAKAS G., KATAKIS I. & VLAHAVAS I. (2011). Random k-labelsets for multilabel classification. *IEEE Transactions on Knowledge and Data Engineering*, **23**(7), 1079–1089.

VINAY J.-P. & DARBELNET J. (1958). *Stylistique comparée du français et de l'anglais : méthode de traduction*. Bibliothèque de stylistique comparée. Didier.

VYAS Y., NIU X. & CARPUAT M. (2018). Identifying Semantic Divergences in Parallel Text without Annotations. In M. A. WALKER, H. JI & A. STENT, Eds., *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, p. 1503–1515 : Association for Computational Linguistics.

ZHAI Y. (2018). Construction d'un corpus multilingue annoté en relations de traduction. In *Rencontre Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*, p. 85–99, Rennes, France.

ZHAI Y., MAX A. & VILNAT A. (2018). Construction of a Multilingual Corpus Annotated with Translation Relations. In *First Workshop on Linguistic Resources for Natural Language Processing*, p. 102–111, Santa Fe, New Mexico, USA.

ZHAI Y., SAFARI P., ILLOUZ G., ALLAUZEN A. & VILNAT A. (2019). Towards Recognizing Phrase Translation Processes : Experiments on English-French. *CoRR*, **abs/1904.12213**.