Post-édition statistique pour l'adaptation aux domaines de spécialité en traduction automatique

Raphaël Rubino, Stéphane Huet, Fabrice Lefèvre, Georges Linarès
LIA-CERI, Université d'Avignon et des Pays de Vaucluse, Avignon, France
{prénom.nom}@univ-avignon.fr

RÉSUMÉ
Cet article présente une approche de post-édition statistique pour adapter aux domaines de spécialité des systèmes de traduction automatique génériques. En utilisant les traductions produites par ces systèmes, alignées avec leur traduction de référence, un modèle de post-édition basé sur un alignement sous-phrastique est construit. Les expériences menées entre le français et l'anglais pour le domaine médical montrent qu'une telle adaptation *a posteriori* est possible. Deux systèmes de traduction statistiques sont étudiés : une implémentation locale état-de-l'art et un outil libre en ligne. Nous proposons aussi une méthode de sélection de phrases à post-éditer permettant d'emblée d'accroître la qualité des traductions et pour laquelle les scores oracles indiquent des gains encore possibles.

ABSTRACT
Statistical Post-Editing of Machine Translation for Domain Adaptation

This paper presents a statistical approach to adapt generic machine translation systems to the medical domain through an unsupervised post-edition step. A statistical post-edition model is built on statistical machine translation outputs aligned with their translation references. Evaluations carried out to translate medical texts from French to English show that a generic machine translation system can be adapted *a posteriori* to a specific domain. Two systems are studied : a state-of-the-art phrase-based implementation and an online publicly available software. Our experiments also indicate that selecting sentences for post-edition leads to significant improvements of translation quality and that more gains are still possible with respect to an oracle measure.

MOTS-CLÉS : Traduction automatique statistique, post-édition, adaptation aux domaines de spécialité.

KEYWORDS: Statistical Machine Translation, Post-editing, Domain Adaptation.

1 Introduction

La traduction automatique statistique basée sur l'alignement sous-phrastique (Koehn *et al.*, 2003) est une approche très populaire qui mène à des performances intéressantes. Les modèles statistiques sous-jacents sont construits à l'aide de phrases en relation de traduction, d'où sont extraites des probabilités d'alignements entre les séquences de mots. Les ressources linguistiques nécessaires à l'estimation de ces probabilités sont les corpus parallèles, éléments indispensables dans le processus de construction du modèle de traduction. Cependant, ces ressources sont coûteuses à

Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 2: TALN, pages 527–534,
Grenoble, 4 au 8 juin 2012.

produire et limitent l'approche. Ce manque de données parallèles se fait ressentir plus fortement pour des domaines spécialisés. En effet, beaucoup d'activités humaines impliquent l'utilisation d'une langue spécifique comportant des particularités syntaxiques et terminologiques (Sager *et al.*, 1980). Ainsi, construire des systèmes de traduction pour tous les domaines semble hors de portée et nous pensons que l'adaptation d'un système « générique » représente une solution viable à la prise en charge des domaines dans leur diversité.

Même dans les cas où la traduction automatique peut atteindre de bonnes performances, il est possible d'améliorer manuellement la sortie des systèmes. Mais cela implique une étape de post-édition qui peut se révéler coûteuse selon l'effort à fournir pour produire une traduction de qualité (Martínez, 2003). Il paraît donc intéressant d'automatiser ce processus d'édition *a posteriori* afin de contrôler et d'améliorer les traductions produites par un système. Nous proposons, dans cet article, d'utiliser une approche de post-édition statistique (*statistical post-edition*, SPE) basée sur l'alignement sous-phrastique afin d'adapter des systèmes de traduction automatique à un domaine de spécialité. Le domaine étudié est celui de la médecine et la paire de langues est français-anglais. Plusieurs systèmes de traductions statistiques sont utilisés et nous proposons différentes méthodes afin d'introduire une petite quantité de données spécialisées pendant le processus de traduction. Nous étudions aussi l'impact de la post-édition en l'appliquant systématiquement à toute traduction, puis en sélectionnant les phrases à post-éditer à l'aide d'un classifieur.

L'organisation de cet article est la suivante. La section 2 présente l'approche de SPE par segments sous-phrastique. Puis, dans la section 3, nous proposons un cadre expérimental et donnons des détails sur les données utilisées ainsi que les différentes configurations évaluées. La section 4 contient les résultats en terme de traduction automatique, suivie de la section 5 présentant les résultats de notre approche pour la post-édition. La sélection des phrases à post-éditer est détaillée dans la section 6.

2 L'adaptation aux domaines par post-édition statistique

La post-édition d'une traduction automatique consiste à générer un texte T'' à partir d'une hypothèse de traduction T' provenant d'un texte source S. Ainsi, ne sont nécessaires à la SPE que des données monolingues dans la langue cible. Le corpus parallèle utilisé pour la construction du modèle de post-édition peut être constitué d'hypothèses de traductions, de leurs références de traduction, d'hypothèses post-éditées manuellement, etc.

Parmi les premiers travaux à relater de l'efficacité de la SPE, Simard *et al.* (2007a) proposent d'éditer automatiquement des hypothèses de traduction produites par un système à base de règles. Une étude détaillée de cette approche (traduction par règles et SPE par alignement sous-phrastique) est proposée par (Dugast *et al.*, 2007) et les gains observés sur la mesure d'évaluation BLEU (Papineni *et al.*, 2002) peuvent atteindre jusqu'à 10 points.

L'amélioration de la qualité des traductions produites par un système à base de règles en utilisant la SPE est donc possible. Certains auteurs y voient la possibilité d'adapter le système de traduction à des domaines de spécialité. Selon (Isabelle *et al.*, 2007; Simard *et al.*, 2007b), en introduisant des données spécialisées lors de la phase de post-édition, il est possible d'adapter un système de traduction *a posteriori*. Cette technique a aussi été appliquée par de Ilarraza *et al.* (2008), à partir de traductions basées sur des règles post-éditées par alignement sous-phrastique, en y ajoutant des informations morphologiques. Parmi ces travaux, il est important de noter que

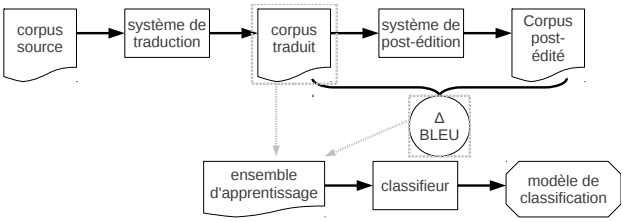


FIGURE 1: Architecture générale combinant la traduction et la SPE, ainsi que la sélection des phrases à post-éditer par estimation des gains $\Delta BLEU$.

certains auteurs ont suggéré de combiner deux systèmes statistiques pour la traduction et la post-édition, sans toutefois poursuivre ces expérimentations (Isabelle *et al.*, 2007; Oflazer et El-Kahlout, 2007).

Plus récemment, Béchara *et al.* (2011) sont les premiers à étudier la combinaison de la traduction et la post-édition par alignement sous-phrastique. Des améliorations sont observées entre le français et l'anglais, en introduisant des informations sur le contexte source dans la phase de post-édition. Mais il n'y a, à notre connaissance, aucune publication sur l'utilisation de l'approche pour l'adaptation aux domaines de spécialité. Nous proposons donc de l'étudier dans cet article, en relation avec la sélection automatique des phrases à post-éditer. Ce dernier aspect est abordé dans (Suzuki, 2011), où les auteurs décrivent une sélection de phrases à post-éditer manuellement basée sur l'estimation de la qualité de traduction au niveau des phrases.

Les travaux présentés dans cet article sont centrés sur l'adaptation de systèmes de traduction génériques en utilisant une petite quantité de données parallèles du domaine. La figure 1 illustre l'architecture générale de notre approche, combinant la traduction, la post-édition et la classification. Dans un premier temps, une traduction automatique d'un texte spécialisé est produite. Puis, en utilisant la référence de traduction, un corpus parallèle monolingue est construit. Ce corpus est utilisé pour construire le modèle de post-édition. Lorsqu'un nouveau texte est traduit, nous proposons alors deux méthodes : la post-édition *naïve* de toutes les phrases ou la sélection des phrases à post-éditer. Cette sélection est faite sur l'estimation des gains possibles par post-édition en recourant à la métrique automatique BLEU.

3 Cadre expérimental

L'idée générale des travaux présentés dans cet article est d'accroître la qualité de traductions de textes spécialisés produites par un système générique par une étape de post-édition. Nous considérons deux types de systèmes de traduction permettant de passer de la langue source à la langue cible : MOSES avec son implémentation de l'alignement sous-phrastique (Koehn *et al.*, 2007) et le système en ligne GOOGLE TRADUCTION[1]. Nous utilisons MOSES pour post-éditer les hypothèses de traduction produites par les deux types de systèmes de traduction individuellement.

Ressources Les données génériques (non spécialisées) sont présentées dans le tableau 1a. Les

1. `http://translate.google.com/`, en utilisant l'API gratuite pendant les mois de Juin et Juillet 2011.

Corpus	Phrases	Mots
Données parallèles		
Europarl v6	1,8 M	50 M
Nations Unies	12 M	300 M
EMEA (Médical)	160 k	4 M
Données monolingues		
News Commentary v6	181 k	4 M
Shuffled News Corpus (2007–2011)	25 M	515 M

(a) Taille des données utilisées en nombre de phrases.

Système		BLEU (%)	p-valeur
MT_g	ML_g	29,9	
MT_g	ML_{g+m}	38,2	0,002
MT_g	ML_m	39,2	0,002
google		44,9	0,002
MT_m	ML_m	46,4	0,007
MT_{g+m}	ML_m	47,2	0,001
MT_{g+m}	ML_{g+m}	47,3	0,75

(b) Scores BLEU obtenus sur le corpus de test du domaine médical.

TABLE 1: Données utilisées (a) et scores BLEU obtenus avec les différents systèmes de traduction (b).

données bilingues, utilisées pour la construction des modèles de traduction, sont composées de la sixième version du corpus Europarl et du corpus issu des Nations Unies. Les données monolingues, utilisées pour la construction des modèles de langue, sont composées des corpus d'actualités *News Crawl* et de la partie langue cible de *News Commentary*. Toutes ces données génériques ont été mises à disposition lors de la campagne d'évaluation *WMT11* [2]. Les données spécialisées sont issues quant à elles de documents provenant de l'agence européenne de médecine (corpus EMEA (Tiedemann, 2009)). Trois catégories médicales sont concernées : des rapports d'évaluation concernant des traitements effectués sur des humains, d'autres effectués sur des animaux, et des documents de médecine générale. Certains documents composant ce corpus sont des prescriptions médicales. Une des caractéristiques de ces documents est une forte redondance au niveau des phrases. Nous décidons de retirer les phrases répétées car elles ne représentent pas un grand intérêt pour la post-édition. En effet, traduire une phrase se trouvant dans l'ensemble d'apprentissage revient a consulter une mémoire de traduction au niveau des phrases. Aussi, nous ne conservons dans le corpus que les phrases d'une longueur inférieure à 80 mots car les alignements obtenus sur des longues phrases sont généralement de qualité moindre. Puis, trois sous-ensembles de phrases sont constituées, formant un corpus d'entraînement (156k phrases), un corpus de développement (ou optimisation, 2k phrases) et un corpus de test (2k phrases).

Systèmes de traduction Parmi les systèmes de traduction utilisés, l'outil accessible en ligne, noté *google* dans les expériences présentées dans cet article, ne peut être modifié. Nous pouvons cependant post-éditer et évaluer les traductions produites par ce système. La boîte à outils de traduction MOSES permet, quant à elle, de construire un modèle de traduction à partir des corpus parallèles et de contrôler chaque étape de ce processus. Ainsi, nous pouvons faire varier les données monolingues et bilingues utilisées.

Pour les modèles de langue (ML), trois modèles 5-grammes sont construits. Un premier est établi sur les données monolingues génériques (ML_g), un second est construit sur la partie langue cible du corpus d'entraînement médical (ML_m). Ces deux modèles sont combinés par interpolation linéaire (ML_{g+m}) selon des poids estimés par le calcul de la perplexité sur le corpus de développement spécialisé. Pour ce dernier modèle, le vocabulaire générique est limité à 1 million de mots les plus fréquents. Le poids optimal associé au modèle de langue spécialisé est de $0,9$ malgré sa petite taille, ce qui montre le niveau de spécificité du domaine médical.

2. `http://www.statmt.org/wmt11/`

Pour les modèles de traduction (MT), MOSES permet de construire une table de traduction et un modèle de réordonnancement en choisissant les données à utiliser. Un premier modèle de traduction est construit avec les données bilingues génériques (MT_g), un second avec les données médicales (MT_m). La combinaison de ces deux modèles permet d'obtenir un modèle mixte (MT_{g+m}). Pour ces trois configurations, les données bilingues sont alignées au niveau des mots avec l'outil MGIZA++ (Gao et Vogel, 2008). Les poids associés aux éléments composant les modèles de traduction sont optimisés sur le corpus de développement médical pour la métrique BLEU selon la méthode *MERT* (Och, 2003).

Afin de construire des systèmes de post-édition pour l'adaptation au domaine médical, nous utilisons les traductions du corpus d'entraînement spécialisé produites par chaque système de traduction individuellement. Chaque sortie est alignée avec la référence de traduction puis utilisée pour construire un modèle de post-édition à l'aide de MOSES (avec les paramètres par défaut). Le corpus de développement spécialisé, une fois traduit par chaque système de traduction, est utilisé afin d'optimiser les poids des composants du modèle de post-édition.

4 Traduction de textes spécialisés

La première série d'expériences porte sur la traduction du corpus de test spécialisé. Les résultats obtenus selon les différentes configurations de systèmes sont présentés dans le tableau 1b. La comparaison entre les systèmes est effectuée selon la méthode d'approximation par sous-échantillonnage aléatoire implémentée dans l'outil FASTMTEVAL (Stroppa *et al.*, 2007). Ces résultats indiquent que la meilleure configuration, avec un score BLEU de 47,3%, est celle combinant les données génériques et spécialisées ($MT_{g+m}ML_{g+m}$). Cependant, ces scores ne sont pas significativement supérieurs à ceux obtenus par $MT_{g+m}ML_m$ (p-valeur=0,75). Nous pouvons donc en conclure que l'intégration des données génériques dans le modèle de langue ne permet pas d'améliorer les performances du système de traduction, ce qui démontre encore une fois la forte spécificité du domaine médical. Cette constatation est plus marquée encore lors de l'utilisation du modèle de traduction générique (MT_g). Introduire des données génériques dans le modèle de langue (MT_gML_{g+m}) dégrade d'un point de BLEU les performances en comparaison avec le score obtenu par MT_gML_m (de 39,2% à 38,2%). Le système de traduction en ligne obtient quant à lui 44,9% de BLEU, soit 1,5 points de moins que le système construit uniquement sur les données spécialisées.

5 Post-édition des traductions

Afin de post-éditer les hypothèses de traduction produites par les systèmes étudiés, le corpus d'entraînement spécialisé est traduit par les différentes configurations et aligné avec sa référence de traduction. Ainsi, un système de post-édition est construit pour chaque système de traduction. Lorsque le corpus de test est traduit, il peut être post-édité dans son intégralité, ou uniquement sur les phrases avec une amélioration possible. Deux scores sont donc calculés : le premier représente l'application *naïve* de la post-édition, le second indique les gains maximums possibles de notre approche (score *oracle*).

Système en ligne Le système de traduction en ligne obtient des résultats convenables lors de la phase de traduction et les résultats obtenus en post-édition sont présentés dans le tableau 2. La comparaison entre les systèmes montre des différences significatives, avec des $p\text{-}valeur$ de 0,001 pour la post-édition *naïve* et 0,05 pour les scores *oracles*. Deux systèmes de post-édition sont

	base	$+ SPE_m ML_m$	$+ SPE_m ML_{g+m}$
google	44,9	46,8 (53,3)	47,9 (53,5)
$MT_g ML_g$	29,9	43,4 (44,2)	45,6 (47,0)
$MT_g ML_m$	39,2	42,7 (44,2)	42,5 (44,4)

TABLE 2: Scores BLEU (%) après post-édition des traductions produites par les systèmes en ligne (*google*) et *Moses* sur le corpus de test médical, utilisant un modèle de langue générique ou spécialisé. Scores *oracle* entre parenthèses.

construits, utilisant les données spécialisées pour le modèle de post-édition et se différenciant selon les modèles de langue, spécialisé ($SPE_m ML_m$) ou mixte ($SPE_m ML_{g+m}$). Les meilleurs résultats sont obtenus par $SPE_m ML_{g+m}$ avec un score BLEU de 47,9%. Le score oracle indique pour cette configuration un score maximum de 53,5%, ce qui motive notre approche de sélection de phrases.

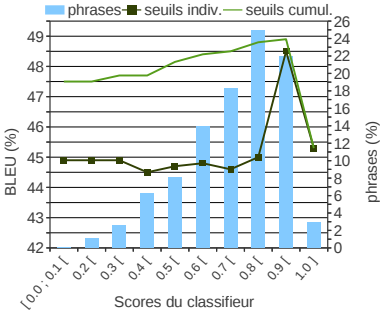
Système générique Nous proposons à présent de post-éditer les sorties d'un système de traduction statistique par alignement sous-phrastique ne disposant pas de données spécialisées pour le modèle de traduction. Les résultats de la post-édition sont présentés dans le tableau 2 en utilisant ($MT_g ML_m$) ou non ($MT_g ML_g$) les données spécialisées pour construire un ML. Le recours à la post-édition montre un gain significatif pour $MT_g ML_g$ en faisant progresser le score BLEU de 29,9% à 45,6% pour la meilleure configuration ($SPE_m ML_{g+m}$). L'augmentation observée de 3,5 points de BLEU à partir de $MT_g ML_m$ est moins importante, tout en restant statistiquement significative ($p\text{-}valeur = 0,001$).

Système spécialisé et mixte Afin de couvrir l'ensemble des expériences possibles selon les différents systèmes de traduction construits, nous procédons à présent à l'évaluation de la post-édition de sorties issues des systèmes utilisant les données spécialisées. Deux systèmes de traduction sont concernés, MT_m et MT_{g+m}, chacun pouvant utiliser un modèle de langue spécialisé ou mixte. Cependant, nous ne détaillons pas les résultats obtenus après post-édition des traductions, car aucune amélioration n'a été observée lors des expérimentations. Seuls les scores *oracles* indiquent qu'un gain est possible, si la sélection des phrases à post-éditer est correctement effectuée.

6 Sélection des phrases à post-éditer

La sélection des traductions à post-éditer est motivée par les scores *oracles* mesurés dans les expériences précédentes. Nous proposons d'utiliser un classifieur afin de détecter les traductions pouvant être améliorées grâce à la post-édition. Le score $\Delta BLEU$ (avant et après post-edition) permet d'associer une classe aux phrases d'un corpus d'entraînement. Nous choisissons le corpus de développement spécialisé comme ensemble d'entraînement pour le classifieur. Ce dernier est de type Séparateur à Vaste Marge (SVM (Boser *et al.*, 1992)), implémenté dans l'outil LIBSVM (Chang et Lin, 2011). Les phrases traduites sont utilisées sous la forme de vecteurs de n-grammes (avec $n \in [1;3]$ dans notre cas).

Nous évaluons notre approche de sélection de phrases à post-éditer en utilisant la configuration dont le score *oracle* est le plus élevé, c-à-d le système de traduction en ligne (*google*) avec le système de post-édition $SPE_m LM_{g+m}$ (*oracle* atteignant 53,5%). Le corpus de développement spécialisé permet de construire le modèle de classification, puis le corpus de test spécialisé est soumis au SVM afin d'en extraire les phrases à post-éditer. Les performances du classifieur



(a) Scores BLEU et nombre de phrases étiquetées « à post-editer » selon les seuils individuels et cumulés des scores issus du classifieur sur le corpus de test médical.

google	initial	+ SPE	+ SPE_selection
TER	42,3	40,4	39,7
BLEU	44,9	47,9	48,9

(b) Scores TER et BLEU après traduction, sélection des phrases et post-édition (p(à post-éditer) $>= 0,8$) sur le corpus de test médical.

FIGURE 2: Analyse sur le corpus de test des résultat de post-édition avec sélection de phrases pour le système de traduction en ligne.

atteignent 79,5% de rappel et 40,1% de précision. Le classifieur produit un score de confiance pour chaque classe attribuée, correspondant à la probabilité qu'une phrase appartienne à la classe prédite. Ceci nous permet de définir des seuils d'acceptation des phrases classées dans la catégorie à post-éditer. Nous évaluons notre approche de sélection selon ces seuils, individuellement ou cumulés. Les résultats sont présentés dans la figure 2a. En cumulant les seuils au dessus de 0,8 pour la classe « à post-éditer », 1 point de BLEU est gagné par rapport à l'application *naïve* de la post-édition. Nous remarquons qu'une quantité plus importante de phrases sont post-éditées entre les seuils 0,5 et 0,8, et seulement 60 phrases le sont avec un score supérieur à 0,9. Cet aspect influence les résultats en terme de score BLEU. L'évaluation globale du corpus de test après post-édition des phrases sélectionnées montre une amélioration selon les deux métriques utilisées (tableau 2b). L'utilisation d'une sélection apparaît donc comme une méthode apportant des gains en comparaison à l'application *naïve* de la post-édition (avec une p-valeur égale à 0,004).

7 Conclusion et perspectives

Nous avons présenté dans cet article une approche de post-édition statistique fondée sur les segments pour l'adaptation aux domaines de spécialité en traduction automatique. Les expériences menées montrent qu'un système de traduction générique peut être adapté *a posteriori* par l'introduction de données spécialisées dans une étape de post-édition. L'application *naïve* de la SPE permet d'améliorer la qualité de traduction dans certains cas. Les scores *oracles* indiquent que pour toutes les configurations étudiées, des gains en terme de score BLEU sont possibles. Les meilleurs résultats sont obtenus par le système de traduction en ligne, couplé avec un système de post-édition construit sur des données mixtes, et en utilisant un classifieur pour sélectionner les traductions à post-éditer. Comparé au système de base, le score BLEU est amélioré de 4 points. L'apprentissage du classifieur est toutefois limité aux hypothèses de traduction et nous envisageons, dans des travaux futurs, d'enrichir les paramètres d'apprentissage pour mieux tenir compte du contexte de traduction afin de s'approcher des scores *oracles* mesurés.

Références

BÉCHARA, H., MA, Y. et van GENABITH, J. (2011). Statistical post-editing for a statistical MT system. *In MT Summit XIII*, pages 308–315.

BOSER, B., GUYON, I. et VAPNIK, V. (1992). A training algorithm for optimal margin classifiers. *In 5th annual workshop on Computational learning theory*, pages 144–152.

CHANG, C.-C. et LIN, C.-J. (2011). LIBSVM : A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27 :1–27 :27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

de ILARRAZA, A., LABAKA, G. et SARASOLA, K. (2008). Statistical postediting : A valuable method in domain adaptation of RBMT systems for less-resourced languages. *In MATMT*, pages 35–40.

DUGAST, L., SENELLART, J. et KOEHN, P. (2007). Statistical post-editing on Systran's rule-based translation system. *In WMT*, pages 220–223.

GAO, Q. et VOGEL, S. (2008). Parallel implementations of word alignment tool. *In ACL Workshop : Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pages 49–57.

ISABELLE, P., GOUTTE, C. et SIMARD, M. (2007). Domain adaptation of MT systems through automatic post-editing. *In MT Summit XI*, pages 255–261.

KOEHN, P., HOANG, H., BIRCH, A., CALLISON-BURCH, C., FEDERICO, M., BERTOLDI, N., COWAN, B., SHEN, W., MORAN, C., ZENS, R. *et al.* (2007). Moses : Open source toolkit for statistical machine translation. *In ACL*, pages 177–180.

KOEHN, P., OCH, F. et MARCU, D. (2003). Statistical phrase-based translation. *In NAACL-HLT*, volume 1, pages 48–54.

MARTÍNEZ, L. (2003). *Human Translation versus Machine Translation and Full Post-Editing of Raw Machine Translation Output*. Thèse de doctorat, Dublin City University.

OCH, F. (2003). Minimum error rate training in statistical machine translation. *In ACL*, volume 1, pages 160–167.

OFLAZER, K. et EL-KAHLOUT, I. (2007). Exploring different representational units in English-to-Turkish statistical machine translation. *In WMT*, pages 25–32.

PAPINENI, K., ROUKOS, S., WARD, T. et ZHU, W. (2002). BLEU : A method for automatic evaluation of machine translation. *In ACL*, pages 311–318.

SAGER, J., DUNGWORTH, D. et MCDONALD, P. (1980). English special languages : principles and practice in science and technology. *Wiesbaden : Oscar BrandstetterVerlay*, pages 2–35.

SIMARD, M., GOUTTE, C. et ISABELLE, P. (2007a). Statistical phrase-based post-editing. *In NAACL-HLT*, pages 508,515.

SIMARD, M., UEFFING, N., ISABELLE, P. et KUHN, R. (2007b). Rule-based translation with statistical phrase-based post-editing. *In WMT*, pages 203–206.

STROPPA, N., OWCZARZAK, K. et WAY, A. (2007). A cluster-based representation for multi-system MT evaluation. *In TMI*, pages 221–230.

SUZUKI, H. (2011). Automatic post-editing based on SMT and its selective application by sentence-level automatic quality evaluation. *In MT Summit XIII*, pages 156–163.

TIEDEMANN, J. (2009). News from OPUS—a collection of multilingual parallel corpora with tools and interfaces. *In RANLP*, volume V, pages 237–248.