

La prédiction avec des règles peu coûteuses, utile pour la TA statistique des forums ?

Johanna Gerlach¹, Victoria Porro¹, Pierrette Bouillon¹, Sabine Lehmann²

(1) UNIVERSITÉ DE GENÈVE FTI/TIM - 40, bvd Du Pont-d'Arve, CH-1211 Genève 4, Suisse

(2) ACROLINX GmbH, Friedrichstr. 100, 10117 Berlin, Allemagne

Johanna.Gerlach@unige.ch, Victoria.Porro@unige.ch,

Pierrette.Bouillon@unige.ch, Sabine.Lehmann@acrolinx.com

RÉSUMÉ

Cet article s'intéresse à la traduction automatique statistique des forums, dans le cadre du projet européen ACCEPT (« Automated Community Content Editing Portal »). Nous montrons qu'il est possible d'écrire des règles de prédiction peu coûteuses sur le plan des ressources linguistiques et applicables sans trop d'effort avec un impact très significatif sur la traduction automatique (TA) statistique, sans avoir à modifier le système de TA. Nous décrivons la méthodologie proposée pour écrire les règles de prédiction et les évaluer, ainsi que les résultats obtenus par type de règles.

ABSTRACT

Can lightweight pre-editing rules improve statistical MT of forum content?

This paper focuses on the statistical machine translation (SMT) of forums within the context of the European Framework ACCEPT (« Automated Community Content Editing Portal ») project. We demonstrate that it is possible to write lightweight pre-editing rules that require few linguistic resources, are relatively easy to apply and have significant impact on SMT without any changes to the machine translation system. We describe methodologies for rule development and evaluation, and provide results obtained for different rule types.

MOTS-CLÉS : prédiction, langage contrôlé, traduction statistique, forums

KEYWORDS : pre-edition, controlled language, statistical machine translation, forums

1 Introduction

Aujourd'hui, les textes communautaires (forums) jouent un rôle de plus en plus important sur le Web. Ils restent cependant difficiles à traduire, en raison de spécificités, qui les rendent plus proches de l'oral que de l'écrit (voir Figure 1).

La sa ne pose pas de problème (du moins on ne recoit pas l'alerte).La dessus on est sur de rien et c'est valable pour n'importe qu'elle antivirus que j'ai put voir, contrairement aux dires.

FIGURE 1 – Exemples tirés de forums informatiques.

Le projet européen ACCEPT (« Automated Community Content Editing PorTal », www.accept.unige.ch) tente de lever ce paradoxe et s'intéresse à trois méthodes pour améliorer la traduction des forums, dont il cherche à mesurer l'impact respectif : la prédiction, la post-édition et des techniques issues de la TA statistique elle-même (par exemple, pour l'adaptation au domaine). Les forums utilisés dans le projet sont ceux des partenaires : Symantec (forums informatiques, fr.community.norton.com) et Traducteurs sans Frontières (textes médicaux). La traduction automatique se fait avec Moses (Koehn et

al., 2007) et la préédition/post-édition avec la plateforme linguistique d'Acrolinx (www.acrolinx.com), l'un des logiciels les plus utilisés aujourd'hui pour le contrôle-qualité de la documentation technique (Bredenkamp et al., 2000). Acrolinx est un logiciel de validation semi-automatique. Pour Accept, il est accessible via un *plug-in* aux membres de la communauté, qui se chargent d'appliquer les règles de préédition et de post-édition, en vue d'améliorer la qualité de la source et de la traduction.

Le travail décrit ici s'inscrit directement dans ce projet et se focalise sur la préédition en français pour les forums de Symantec : est-il possible d'écrire avec la plate-forme d'Acrolinx des règles de préédition utiles pour la TA statistique français-anglais et avec quel impact ? Notre but final est double : décrire les règles utiles ; mesurer ensuite s'il est possible d'obtenir le même gain avec d'autres méthodes (Cf. Rayner et al., 2012). Dans cet article, nous décrivons d'abord la méthodologie utilisée pour définir les règles (Section 2). Nous présentons ensuite les règles développées (Section 3). La dernière partie sera consacrée à l'évaluation de l'impact des différentes règles sur la traduction (Sections 4 et 5). Un problème est d'évaluer les règles de manière rapide et fiable, sans référence : nous comparons les résultats obtenus avec des traducteurs et des juges recrutés avec Amazon Mechanical Turk (AMT).

2 La préédition

La préédition revêt des réalités très différentes en traduction automatique (TA) : correction orthographique et grammaticale ; normalisation lexicale du texte source (par exemple, Han et Baldwin, 2011, Banerjee et al., 2012) ; langage contrôlé (O'Brien, 2003) ; règles de réordonnement (par ex., Wang et al., 2007, Genzel, 2010). De manière générale, peu d'outils de préédition s'intéressent à ces différents types de préédition en même temps. Pour des raisons en partie historiques, le langage contrôlé a été d'avantage associé à la TA linguistique, par règles (Pym, 1988, Bernth et Gdaniec, 2002, O'Brien et Roturier, 2007, etc.) (à l'exception de Aikawa et al. 2007) ; en revanche, la correction orthographique, la normalisation lexicale et les règles de réordonnement ont toujours fait partie intégrante de la TA statistique. Dans ce travail, dans une optique plus éclectique, nous avons développé des règles des quatre types vus plus haut, qui répondent aux critères suivants :

- Elles se focalisent sur quatre phénomènes qui ont un impact clair sur la TA statistique des forums, à savoir les problèmes de confusion entre mots (liés aux homophones), la langue informelle et familière (généralement absente des données d'entraînement), la ponctuation et les différences syntaxiques entre le français et l'anglais.
- Les fautes doivent pouvoir être détectées avec les règles d'Acrolinx. Celles-ci sont décrites avec un langage de patrons, qui repose sur un étiquetage syntaxique des textes (Bredenkamp, 2000). Ceci a évidemment des conséquences sur le type de règles développées : il est difficile de détecter/corriger avec précision les fautes non-locales. Par contre, les règles sont facilement portables dans d'autres outils puisqu'elles nécessitent très peu de ressources linguistiques.
- Les premiers tests à Symantec ont montré que la communauté des utilisateurs des forums Symantec ne semble pas disposée à passer beaucoup de temps sur la préédition. La précision est donc plus importante que le rappel et il est important que l'outil de préédition produise des suggestions de corrections, si possible uniques.

A défaut de données post-éditées qui permettraient d'identifier automatiquement les phénomènes mal traduits, les règles ont toutes été définies manuellement avec Acrolinx, qui offre une plate-forme complète pour développer, déboguer et tester les règles sur des corpus (Bredenkamp et al., 2000). Développer une règle implique de passer par les étapes suivantes (Figure 2) : 1) identifier une règle a priori utile pour la TA statistique, par exemple 'éviter « si... et que... »' ; 2) définir un ou plusieurs patrons (« Trigger ») correspondants sous forme d'expression régulière pour identifier le phénomène dans les textes; 3) proposer une transformation (« Suggestion ») plus traduisible, ici remplacer « que » par « si » pour obtenir « si... et si... »; 4) appliquer la règle sur le corpus de test et traduire les phrases prééditées et non prééditées de manière à produire le fichier-résultat de la Figure 3; 5) vérifier les résultats et ajouter si nécessaire des exceptions pour bloquer la règle dans certains cas (« *Negative evidence* »). La Figure 2 résume la règle « si... et ... que ».

Patron (Trigger) : @conj [2-15] 'et' 'que'	Conjonction « si » suivie de 2 à 15 mots puis de la conjonction « et » et de « que »
Suggestion : 'que' -> @conj	« que » est remplacé par la conjonction « si »
Exception (<i>Negative evidence</i>) : que [+@conj]* que	Bloquer la règle si la conjonction « si » est précédée par « que » + un ou plusieurs mots.

FIGURE 2 – Règle Acrolinx « si et ... que ... »

Source1	Source2	Translation1	Translation2
Si ton problème est résolu et que tout marche bien, pense à supprimer tes points de restauration.	Si ton problème est résolu et si tout marche bien, pense à supprimer tes points de restauration.	If your problem is solved and that everything is working well, think to remove your restore points.	If your problem is solved and if all goes well, think to remove your restore points.

FIGURE 3 – Extrait du fichier résultat

Deux ressources se sont révélées particulièrement utiles pour alimenter les règles : les mots inconnus du système statistique (OOV), extraits des données-test avec Moses sur la base des corpus-test, qui sont des bons indicateurs de ce qui n'est pas couvert par les données d'entraînement (voir aussi Banerjee et al., 2012) et des listes des bigrammes/trigrammes fréquents dans le corpus de test, mais absents de celui d'entraînement, avec la traduction des phrases correspondantes, qui sont souvent des segments mal traduits. Dans la suite, nous décrivons l'ensemble de règles développé suivant cette méthodologie.

3 Les règles développées

Notre but est donc de développer des règles utiles pour la TA statistique qui suivent les critères définis dans la Section 2. Celles-ci peuvent être classées en fonction de différentes dimensions, dont nous mesurerons l'impact dans la suite : règles pour les humains qui améliorent le texte source ou règles pour la TA uniquement ; règles automatiques ou manuelles ; catégorie de règles ; règles avec une ou plusieurs suggestions ou sans suggestion. Acrolinx permet de définir aisément des ensembles de règles et des ordres d'application différents. Pour Symantec, les règles ont été regroupées en trois ensembles, destinés à être utilisés en séquence dans leur *plug-in* de prédiction (www.accept-portal.eu, Roturier et al., 2012) et à réduire le plus possible le rôle des utilisateurs :

- Le premier ensemble (**Ensemble 1**) comprend les règles automatiques, qui doivent être appliquées en premier lieu, pour limiter le bruit dans les autres ensembles (2 et 3). Il inclut la plupart des règles pour la confusion de mots de différentes catégories syntaxiques, liée aux homophones, et gère également la ponctuation et l'élision, par exemple : *Merci beaucoup je fais sa de suite* → *Merci beaucoup, je fais ça de suite*.
- Le deuxième (**Ensemble 2**) comprend les règles avec plusieurs suggestions ou sans suggestion où l'utilisateur doit nécessairement intervenir pour le contrôle. Cet ensemble inclut les règles de grammaire pour l'accord et la confusion des temps/modes, ainsi que les règles de style, en particulier pour éviter le langage familier et informel (questions directes, phrases clivées, mots familiers, troncations, etc.), par exemple : *Tu as lu le tuto sur le forum?* → *As-tu lu le tutoriel sur le forum ?*
- Finalement, le troisième (**Ensemble 3**) regroupe les règles automatiques pour la TA, qui n'améliorent pas nécessairement la qualité du texte source. Celles-ci modifient l'ordre des mots pour les rendre plus proches de l'anglais ou pour éviter des ambiguïtés (*Je te le donne en pièce jointe* → *je te donne ça en pièce jointe* ; *j'ai tout pris* → *j'ai pris tout*) ou encore transforment des mots ou expressions mal traduits dans un équivalent plus traduisible (« ne » ... « que » → *uniquement*, « soit » ... « soit » → *ou*, etc.). Une des règles automatiques convertit la deuxième personne du singulier informelle dans le correspondant formel, beaucoup plus fréquent dans les données d'entraînement (Rayner et al., 2012) : *As-tu lu le tutoriel sur le forum?* → *Avez-vous lu le tutoriel sur le forum ?*

La suite se focalise sur l'évaluation de ces différentes règles sur des textes extraits des forums de Symantec. Nous décrivons d'abord la méthodologie de l'évaluation, puis discutons les résultats.

4 Méthodologie de l'évaluation

4.1 Sélection des données

Afin de constituer un corpus représentatif, nous avons sélectionné 10 000 phrases des données fournies par Symantec, sur la base des mots et des bigrammes fréquents dans l'ensemble des données, en conservant les mêmes proportions de phrases de chaque longueur. Pour simuler l'utilisation des règles décrite en 3, les trois ensembles de règles ont été appliqués en séquence, chacun prenant en entrée le corpus entier avec les corrections de l'étape précédente.

Pour chaque ensemble, les règles ont donc été appliquées une par une et les phrases corrigées, en suivant les suggestions de correction proposées par nos règles. Toujours afin de simuler l'utilisation prévue, cette étape diffère selon les ensembles de règles : pour les premier et troisième ensembles, les corrections ont été appliquées automatiquement, sans vérification de la suggestion ; pour le deuxième, manuellement. Par conséquent, pour les ensembles un et trois, la correction a pu produire des erreurs, vu que la précision des règles n'est pas parfaite.

Finalement, les deux sources, brutes et prééditées, ont été traduites en anglais avec le système de TA statistique développé avec Moses dans le cadre du projet ACCEPT (Accept Deliverable D4.1, 2012). Celui-ci a été entraîné avec les données d'Europarl

(<http://www.statmt.org/wmt12/>), ainsi que des manuels techniques de Symantec. A titre indicatif, le score Bleu est de 42.41 sur un extrait de 500 phrases. Pour l’évaluation, 50 phrases maximum par règle ont été retenues (soit un total de 1 733 phrases), d’où ont été extraites toutes les phrases où les cibles étaient différentes (soit un total de 1 364 phrases).

4.2 Annotation

Pour évaluer l’impact des règles sur la traduction, nous avons opté pour une évaluation humaine comparative de la traduction des phrases brutes et prééditées. Nous présentons donc aux évaluateurs des groupes de phrases du type {*source*, *traduction_1* brute, *traduction_2* de la phrase prééditée}, où *traduction_1* et *traduction_2* apparaissent dans un ordre aléatoire, avec les différences marquées en couleur (Figure 4). Pour les premier et deuxième ensembles de règles, la phrase source correspond à la phrase prééditée. Pour le troisième, il s’agit de la source non prééditée, puisque le résultat de la modification pourrait ne pas être du français correct, par exemple suite à un réordonnement. Les évaluateurs doivent ensuite choisir l’un des cinq jugements comparatifs : « *first clearly better* », « *first slightly better* », « *about the same* », « *second slightly better* », « *second clearly better* ».

Nous avons eu recours à deux types d’évaluateurs, afin de comparer les résultats : d’une part, des travailleurs recrutés sur Amazon Mechanical Turk (AMT), d’autre part des traducteurs de langue maternelle anglaise en fin de formation à la Faculté de Traduction et d’Interprétation (FTI) de l’Université de Genève. Comme dans une précédente étude (Rayner et al., 2012), nous avons imposé les restrictions suivantes aux travailleurs AMT : 1) qu’ils soient de langue maternelle anglaise et qu’ils résident au Canada, afin d’augmenter la probabilité qu’ils soient bilingues anglais-français, et 2) qu’ils aient un historique de travail fiable sur AMT. Pour les évaluateurs ‘traducteurs’, qui n’ont pas accès à la plate-forme AMT, actuellement limitée aux résidents des Etats-Unis/Canada, nous avons développé une application Windows afin qu’ils puissent effectuer l’évaluation dans des conditions similaires que sur AMT. Tous les évaluateurs ont été payés pour la tâche.

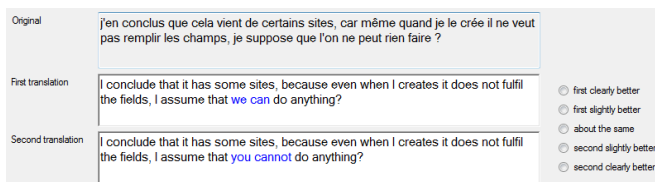


FIGURE 4 – Interface d’évaluation dans notre application

Nous avons ainsi récolté 6 jugements pour chaque groupe de phrases, 3 par des travailleurs AMT, 3 par des traducteurs. Les résultats seront discutés dans la section suivante.

5 Résultats

5.1. Résultats par ensemble et par type de juges

Le Tableau 1 présente les résultats globaux pour les trois ensembles de règles et les deux types de juges (AMT, traducteurs). Pour chaque ensemble, nous avons calculé le pourcentage d’application des règles sur 10000 phrases et, pour les données d’évaluation (1733 phrases), la précision (nombre de suggestions correctes sur l’ensemble des suggestions) et l’impact

des règles sur la traduction. Ce dernier a été mesuré en regroupant les catégories *first/second* «*slightly better*» et «*clearly better*» pour obtenir les deux catégories «*Raw better/Pre-edited better*» et en gardant le jugement majoritaire pour chacun des groupes de juges. Le caractère significatif des résultats a été calculé avec le test de McNemar, en comparant les résultats des catégories «*Raw better*» et «*Pre-edited better*».

	% application	Inclus dans l'évaluation	NoImpact	%	Raw Better	%	About hte same	%	Pre-edited Better	%	No majority judgement	%	impact	significatif p<0,05
Ensemble 1 (précision = 91%)														
AMT	42%	611	183	30%	59	10%	50	8%	297	49%	22	4%	pos	oui
trad.	42%	611	183	30%	56	9%	85	14%	258	42%	29	5%	pos	oui
Ensemble 2 (précision = 88%)														
AMT	20%	674	103	15%	114	17%	37	6%	393	58%	27	4%	pos	oui
trad.	20%	674	103	15%	109	16%	73	11%	360	53%	29	4%	pos	oui
Ensemble 3 (précision = 98%)														
AMT	36%	448	83	19%	77	17%	28	6%	239	53%	21	5%	pos	oui
trad.	36%	448	83	19%	69	15%	53	12%	224	50%	19	4%	pos	oui

TABLEAU 1 – Résultats par ensemble et par type de juges

Les résultats montrent qu'il est possible d'arriver à une précision élevée (entre 98% pour l'Ensemble 3 et 88% pour l'Ensemble 2) et que les trois ensembles de règles ont un impact statistiquement significatif sur la traduction ($p < 0,05$), même s'ils ont été appliqués en séquence et partiellement automatiquement. Ces résultats sont très proches pour les trois ensembles de règles et les deux types de juges (AMT/traducteurs). Une comparaison plus approfondie des résultats avec les juges AMT et traducteurs montre cependant que ces derniers ont été plus rapides et arrivent plus souvent à un jugement majoritaire ou unanime (Tableau 2). Les deux groupes convergent vers le même jugement majoritaire dans 75% des cas. Le coefficient de concordance kappa (calculé sur la base des jugements majoritaires pour chacun des deux groupes de juges) est de 0,53.

Evaluateurs	Temps moyen pour l'évaluation de 20 phrases (minutes)	Accord observé
AMT	06 :13	82%
traducteurs	04 :09	89%

TABLEAU 2 – Temps et accord entre types de juge

Notre conclusion est donc qu'il est tout à fait possible de recourir à des juges AMT pour ce type de tâche assez simple, mais que les traducteurs restent plus efficaces.

5.2. Résultats par catégorie d'erreurs

Nous avons ensuite mesuré l'impact sur la traduction par type de règles, en prenant cette fois-ci le jugement majoritaire pour les 6 juges ensemble. Les catégories retenues sont, pour les **Ensembles 1 et 2**, 1) **punctuation** (y compris les règles pour l'éllision), 2) **grammaire (accord)**, 3) **grammaire (autres)**, avec les règles qui régissent l'utilisation des temps/modes, 4) **homophones** (avec toutes les règles qui traitent les confusions au niveau des catégories syntaxiques), 5) style **informel** et pour **l'Ensemble 3**, 1) **clitiques** (avec toutes les règles de réordonnement/reformulation spécifiques aux clitiques), 2) **reformulation** (remplacement d'une expression par une autre), 3) **ordre** des mots (pour les phénomènes

autres que les clitiques) et 4) **tu-vous** (règle qui remplace la deuxième personne informelle par la deuxième personne formelle) (cf. Section 3).

Le Tableau 3 présente les résultats sur 10 000 phrases, regroupés par catégorie, triés d'après le nombre de cas marqués. Nous voyons que nous obtenons une amélioration significative de la traduction pour toutes les catégories, sauf pour **grammaire (autres)**. Cette dernière a un impact négatif car elle tend à générer des données non couvertes par le système (avec un subjonctif ou conditionnel, par exemple, à la place de l'indicatif). Parmi les catégories à impact positif, les plus significatives concernent la ponctuation, les confusions de mots dues aux homophones et à la langue informelle; la moins significative est la catégorie **ordre**, comme nous l'avons déjà constaté sur des données antérieures (Accept Deliverable D2.1, 2012) : les règles de réordonnement s'avèrent en effet assez fragiles, si elles ne sont pas également appliquées aux données d'entraînement.

Catégorie	Total de cas marqués	Inclus dans l'évaluation	NoImpact	%	Raw Better	%	About hte same	%	Pre-edited Better	%	No majority judgement	%	p-value	significatif
ponctuation	3796	416	147	35%	33	8%	32	8%	184	44%	20	5%	2.4E-24	oui
tu	1968	50	20	40%	3	6%	4	8%	21	42%	2	4%	5.2E-04	oui
clitiques	1206	150	32	21%	27	18%	14	9%	69	46%	8	5%	2.9E-05	oui
informel	971	367	42	11%	67	18%	19	5%	216	59%	23	6%	1.4E-18	oui
homophones	659	323	55	17%	38	12%	33	10%	185	57%	12	4%	1.4E-22	oui
grammaire (accord)	591	150	32	21%	22	15%	11	7%	82	55%	3	2%	7.2E-09	oui
reformulation	177	177	19	11%	26	15%	8	5%	115	65%	9	5%	1.3E-13	oui
ordre	71	71	12	17%	17	24%	3	4%	34	48%	5	7%	2.5E-02	oui
grammaire (autres)	36	28	9	32%	9	32%	2	7%	7	25%	1	4%	8.0E-01	no n

TABLEAU 3 – Résultats par catégorie

6 Conclusion

Dans cet article, nous avons montré qu'il est possible d'écrire, pour un domaine, des règles automatiques de prédiction peu coûteuses sur le plan des ressources linguistiques et applicables sans trop d'effort avec un impact très significatif sur la TA statistique, sans avoir à modifier le système statistique. L'évaluation peut se faire avec des juges AMT, qui conviennent bien pour la tâche d'évaluation proposée ici.

L'étape suivante sera de montrer l'impact de ces règles appliquées ensemble et de voir comment les règles sont utilisées par la communauté Symantec. Nous voulons également vérifier si l'effort de post-édition diminue avec les règles de prédiction (Aikawa et al. 2007). Comme mentionné dans l'introduction, l'un des objectifs du projet ACCEPT est de voir s'il est possible d'obtenir le même gain avec d'autres méthodes. Une étude faite en parallèle (Rayner et al., 2012) a montré qu'il est plus efficace d'appliquer la règle « tu-vous » qui transforme la deuxième personne du singulier informelle en deuxième personne du singulier formelle lors de la prédiction que de générer, avec la même règle inversée, des données d'entraînement avec la deuxième personne informelle. Nous comptons tester de la même manière les autres types de règles.

Références

Accept Deliverable D4.1 (2012), <http://www.accept.unige.ch/Products/>

Accept Deliverable D2.1 (2012), <http://www.accept.unige.ch/Products/>

AIKAWA, T., SCHWARTZ, L., KING, R., CORSTON-OLIVER, M., & LOZANO, C. (2007). Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. *In Proceedings of MT Summit XI*, Copenhagen, Denmark.

BANERJEE, P., NASKAR, ROTURIER, J., WAY, A. & VAN GENABITH J. (2012). Domain Adaptation in SMT of User-Generated Forum Content Guided by OOV Word Reduction: Normalization and/or Supplementary Data? *In Proceedings of EAMT*, Trento.

BERNTH, A., & GDANIEC, C. (2002). MTranslatibility. *In Machine Translation* 16, pages 175-218.

BREDEKAMP, A., CRYSMANN, B., & PETREA, M. (2000). Looking for Errors : A Declarative Formalism for Resource-Adaptive Language Checking. *In Proceedings of LREC 2000*. Athens, Greece.

GENZEL, D. (2010). Automatically learning source-side reordering rules for large scale machine translation. *In Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China.

HAN, B. & BALDWIN, T (2011). Lexical Normalisation of Short Text Messages: Makn *Sens a #twitter*. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

KOEHN, P. ET AL. (2007). Moses: open source toolkit for statistical machine translation. *In ACL-2007: Proceedings of demo and poster sessions*, Prague, Czech Republic, pp.177–180.

O'BRIEN, SH. (2003). Controlling controlled English: An Analysis of Several Controlled Language Rule Sets. *In EAMT-CLAW-03, Dublin City University*, pages 105-114.

O'BRIEN, SH. & ROTURIER, J. (2007). How Portable are Controlled Languages Rules? A Comparison of Two Empirical MT Studies. *In MT Summit XI*, Copenhagen, pages 105-114.

PYM, P. J. (1988). Pre-editing and the use of simplified writing for MT: an engineer's experience of operating an MT system. *In Translating and the Computer* 10.

RAYNER, M., BOUILLON, P. & HADDOW, B. (2012). Using Source-Language Transformations to Address Register Mismatches in SMT . *In Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA)*, San Diego, USA.

ROTURIER, J., MITCHELL, L., GRABOWSKI, R., SIEGEL, M. (2012). Using Automatic Machine Translation Metrics to Analyze the Impact of Source Reformulations. *In Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA), October 2012, San Diego, USA*.

WANG, CH., COLLINS, M. & KOEHN, PH., Chinese Syntactic Reordering for Statistical Machine Translation. *In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 737-745.