

## Prise en compte de dépendances syntaxiques pour la traduction contextuelle de segments

Aurélien Max<sup>1</sup>, Rafik Makhloufi<sup>2</sup> et Philippe Langlais<sup>3</sup>

aurelien.max@limsi.fr

rafik.makhloufi@utt.fr

felipe@iro.umontreal.ca

(1) LIMSI-CNRS et Université Paris-Sud 11, Orsay, France

(2) Université de Technologie de Troyes, France\*

(3) DIRO, Université de Montréal, Canada

**Résumé.** Dans un système standard de traduction statistique basé sur les segments, le score attribué aux différentes traductions d'un segment ne dépend pas du contexte dans lequel il apparaît. Plusieurs travaux récents tendent à montrer l'intérêt de prendre en compte le contexte source lors de la traduction, mais ces études portent sur des systèmes traduisant vers l'anglais, une langue faiblement fléchiée. Dans cet article, nous décrivons nos expériences sur la prise en compte du contexte source dans un système statistique traduisant de l'anglais vers le français, basé sur l'approche proposée par Stroppa *et al.* (2007). Nous étudions l'impact de différents types d'indices capturant l'information contextuelle, dont des dépendances syntaxiques typées. Si les mesures automatiques d'évaluation de la qualité d'une traduction ne révèlent pas de gains significatifs de notre système par rapport à un système à l'état de l'art ne faisant pas usage du contexte, une évaluation manuelle conduite sur 100 phrases choisies aléatoirement est en faveur de notre système. Cette évaluation fait également ressortir que la prise en compte de certaines dépendances syntaxiques est bénéfique à notre système.

**Abstract.** In standard phrase-based Statistical Machine Translation (PBSMT) systems, the score associated with each translation of a phrase does not depend on its context. While several works have shown the potential gain of exploiting source context, they all considered English, a morphologically poor language, as the target language. In this article, we describe experiments on exploiting the source context in an English→French PBSMT system, inspired by the work of Stroppa *et al.* (2007). We report a study on the impact of various types of features that capture contextual information, including syntactic dependencies. While automatic metrics do not show significative gains relative to a baseline system, a manual evaluation of 100 randomly selected sentences concludes that our context-aware system performs consistently better. This evaluation also shows that some types of syntactic dependencies can participate to the gains observed.

**Mots-clés :** Traduction automatique statistique, contexte source, dépendances syntaxiques.

**Keywords:** Statistical Machine Translation, source context, syntactic dependencies.

---

\* Ce travail a été effectué lors d'un stage au LIMSI-CNRS.

# 1 Introduction

Les approches statistiques à la traduction automatique reposent en général sur des analyses relativement surfaciques du texte à traduire. En particulier, l'approche basée sur les segments (*Phrase-based Statistical Machine Translation* (PBSMT)) (Koehn *et al.*, 2003) utilise comme unités de traduction des segments constitués de mots contigus sans substrat linguistique, mais qui de fait, représentent un contexte local, et qui mènent à des gains significatifs relativement à l'approche statistique basée sur les mots. Cependant, les systèmes PBSMT standard n'exploitent pas le contexte d'un segment lors de sa traduction. Cela ne permet pas en particulier de sélectionner la traduction appropriée pour des mots polysémiques, autrement que par le fait que la traduction correcte serait la plus fréquente ou qu'elle serait préférée par le modèle de langue cible. Dans l'exemple suivant, le verbe polysémique anglais *to save* apparaît dans trois contextes menant à des traductions différentes en français:

*Follow the instructions outlined below to **save** that file.* → sauvegarder  
*Quitting smoking is a sure-fire way to **save** some money.* → économiser  
*Brown's gamble may **save** the banks but the economy cannot wait.* → sauver

Avec des données d'apprentissage suffisantes, de bonnes traductions pourraient être apprises pour des segments tels que *save that file*, *save some money* et *save the banks*. Toutefois, les éléments nécessaires à la désambiguïsation de sens sont souvent hors de portée de l'approche basée sur les segments, comme illustré dans l'exemple suivant, où il est peu probable que des traductions fiables soient apprises pour le segment souligné:

*The creation of a UK 'bad bank' will be the best way to save troubled UK banks.*

Après le relatif échec des tentatives d'intégration de systèmes de désambiguïsation lexicale dans des systèmes statistiques (Carpuat & Wu, 2005), des travaux récents tels que (Stroppa *et al.*, 2007; Carpuat & Wu, 2007; Gimpel & Smith, 2008) ont proposé de rendre les systèmes PBSMT sensibles au contexte des segments à traduire en réestimant pour chaque segment la probabilité de ses traductions en fonction de son contexte. Les principales informations utilisées dans ces études sont issues du contexte immédiat des segments (par exemple les mots encadrant le segment à traduire). Si quelques améliorations de la qualité des traductions ont été obtenues, il apparaît nécessaire de considérer des contextes plus riches pour améliorer la performance de ce type de désambiguïsation.

Une solution simple consiste à entraîner des systèmes avec des données proches de celles à traduire, ce qui a notamment pour effet de diminuer la polysémie en langue source. Langlais *et al.* (2006) rapportent des gains importants dans une expérience où ils adaptent un système à traduire des textes du domaine médical. Une autre approche consiste à réestimer les probabilités de traduction en fonction de la présence de mots *déclencheurs* dans un contexte un peu plus large (Lavecchia *et al.*, 2008), qui peut par exemple être celui de la phrase du segment à traduire. Cependant, de telles cooccurrences ne sont pas suffisantes pour désambiguïser de façon fiable car elles mettent en jeu des mots qui ne sont pas nécessairement liés dans tous les contextes, comme illustré dans l'exemple suivant:

*Businesses can make deposits right from the office and **save** trips to the bank.* → \*sauver

La prise en compte du contexte pour désambiguïser la traduction d'un segment doit donc pouvoir se baser en partie sur des relations entre mots. Dans cet article, nous présentons une approche qui prend en compte les dépendances syntaxiques liant les mots capturés par des seg-

ments et des mots en dehors de ces segments pour participer à la réestimation des probabilités de leurs traductions. Nous décrivons tout d'abord les travaux existant dans le domaine (section 2), puis proposons une approche basée sur une classification basée sur la mémoire à partir d'exemples exploitant différents traits dont des dépendances syntaxiques (section 3). Nous décrivons alors nos expériences et analysons les résultats obtenus (section 4), puis nous présentons finalement nos travaux futurs et concluons (section 5).

## 2 Travaux sur la prise en compte du contexte source

L'analyse fine des sorties d'un système de traduction PBSMT menée par Vilar *et al.* (2006) met en évidence la nécessité de prendre en compte le contexte source des segments à traduire: les erreurs portant sur les sens des mots traduits représentent 21,9% des erreurs en traduction de l'anglais vers l'espagnol et 28,2% de l'espagnol vers l'anglais. En outre, les formes incorrectes des mots traduits, dues notamment à des absences d'accords, représentent respectivement 33,9% et 9,9%, ce qui peut être imputé à l'absence de transmission de dépendances syntaxiques telles que *sujet-verbe* ou *modifieur-nom* entre mots sources et cibles.

Plusieurs travaux récents en traduction PBSMT ont porté sur l'estimation de probabilités de traduction dépendant du contexte source. Carpuat & Wu (2007) intègrent un système de désambiguïsation lexicale existant qui utilise un classifieur par segment source. Ce classifieur exploite des collocations locales, des mots du contexte immédiat des segments, leur catégorie morphosyntaxique ainsi que des sacs-de-mots et des dépendances simples<sup>1</sup>. Une approche similaire est étudiée dans (Giménez & Márquez, 2007).

Stroppa *et al.* (2007) proposent d'entraîner un classifieur global dont la sortie correspond à des classes pondérées correspondant aux traductions possibles d'un segment source dans son contexte. Les traits utilisés incluent les séquences de mots et de catégories morphosyntaxiques d'un segment source, ainsi que les mots et catégories du contexte immédiat.

L'étude la plus poussée dans cette lignée est celle décrite par Gimpel & Smith (2008), où chaque information contextuelle donne lieu à un score utilisé par le décodeur. Les traits considérés par les auteurs reprennent ceux de (Stroppa *et al.*, 2007), mais incluent également des indices syntaxiques dérivés d'une analyse en constituants (ex: un segment correspond-t-il à un constituant? Un segment contient-il la tête d'un constituant? Nature du non-terminal le plus profond dans le constituant couvrant le segment, etc.), des traits sur la position du segment dans la phrase (source), ainsi que des traits sur la taille du segment et de la phrase.

Si ces travaux ont permis d'obtenir des gains avec les métriques automatiques de qualité des traductions, ces gains concernent des systèmes ayant l'anglais en langue cible, langue morphologiquement pauvre pour laquelle les accords entre mots ont peu d'impact sur la qualité générale des traductions. Gimpel & Smith (2008) n'obtiennent pas de gains en traduction de l'anglais vers l'allemand, ce qui peut être en partie imputé à la qualité du parseur syntaxique utilisé. Allauzen *et al.* (2009) observent avec une approche comparable exploitant des traits lexicaux et morphosyntaxiques des gains modestes dans une tâche de traduction de l'anglais vers le français.

---

<sup>1</sup>Ces traits ne sont pas décrits plus avant par les auteurs.

## 3 Prise en compte du contexte dans un système PBSMT

### 3.1 Estimation de scores de traduction contextuels

Un système de traduction PBSMT produit une traduction en maximisant une fonction de score censée mesurer la qualité d'une traduction candidate étant donnée une phrase à traduire. Typiquement, cette fonction prend la forme d'une combinaison pondérée (soient  $\lambda$  les poids) de traits ou modèles  $h_m$  incluant un modèle de langue ( $l_m$ ) et un ou plusieurs modèles de traduction:

$$P(e|f) \propto \sum_k \sum_m \lambda_m h_m(f_k, e_k) + \lambda_{l_m} P(e)$$

où les traits peuvent être considérés comme des approximations de  $P(e_k|f_k)$ , la traduction d'un segment cible étant donné un segment source. Cette décomposition fait ressortir le fait que les modèles font l'hypothèse que les traductions des différents segments sont des événements indépendants. L'intégration de l'information contextuelle est réalisée dans cette étude en autorisant la prise en compte de traits de la forme:  $h(f_k, e_k, C(f_k, f))$  où  $C(f_k, f)$  désigne le contexte du segment  $f_k$  dans la phrase  $f$ .

Dans (Gimpel & Smith, 2008), un score par type d'élément de contexte est utilisé, ce qui rend les modèles difficiles à estimer de manière fiable. Cette difficulté peut néanmoins être compensée par l'ajustement adéquat des poids de la combinaison linéaire réalisée sur un corpus de développement, ce qui peut s'avérer difficile en pratique lorsque le nombre de poids à ajuster est grand. À l'inverse, Stroppa *et al.* (2007) capturent tous les éléments du contexte dans un unique classifieur qui utilise un arbre de décision. Celui-ci associe à une entrée  $\langle f_k, C(f_k, f) \rangle$  un ensemble pondéré d'étiquettes de classes qui correspondent aux traductions possibles de  $f_k$  dans son contexte.

La collection d'exemples se fait en parcourant l'ensemble des bi-segments obtenus par alignement du bitexte d'entraînement et en extrayant pour chaque couple les valeurs des traits contextuels. La classification opérée réalise un lissage de manière implicite en basant les poids des classes de sortie sur un vote correspondant à la quantité des traits pour lesquels il y a correspondance. En cas de correspondance exacte entre les contextes d'un segment à traduire et d'un segment vu lors de l'apprentissage, le segment cible correspond à la classe retournée de meilleur score, ce qui confère à l'approche des propriétés comparables aux systèmes de traduction basés sur des exemples (*Example-based Machine Translation (EBMT)*).

Le score attribué par le classifieur à la traduction d'un segment en contexte est intégré dans le décodeur en l'ajoutant à la combinaison de modèles. Nous ajoutons également un score binaire qui vaut 1 pour la traduction la plus vraisemblable selon le classifieur et 0 pour les autres traductions. L'importance de ces deux scores est ajustée via leur coefficient de pondération respectif qui est ajusté en même temps que les autres à l'aide d'un corpus de développement.

### 3.2 Sélection de traits contextuels

Outre la séquence des mots constituant un segment, la séquence de leur catégorie morphosyntaxique peut participer à la désambiguïsation des traductions de segments. Les contraintes exprimées sont avant tout grammaticales, ce qui permet en particulier de distinguer des homographes, comme dans l'exemple suivant entre un nom et un verbe:

*The goalkeeper was given the honour down to the fantastic **save** he made → arrêt*  
*Quitting smoking is a sure-fire way to **save** some money. → économiser*

De même, les mots encadrant un segment et leur catégorie morphosyntaxique peuvent aider à désambiguïser entre homographes et dans une certaine mesure entre segments impliquant des sens distincts, notamment en fonction de la portée considérée pour définir le contexte immédiat. Cependant, certaines désambiguïisations ne pourront être réalisées qu'en exploitant des dépendances à plus longue portée, telle que des relations syntaxiques. La traduction correcte du verbe *save* dans cet exemple ne pourra ainsi se faire qu'en considérant son objet direct<sup>2</sup>:

*The creation of a UK 'bad bank' will be the best way to **save** troubled UK banks.*

Nous proposons donc d'extraire les informations concernant les dépendances syntaxiques impliquant un mot appartenant au segment considéré et un mot à l'extérieur. Dans la suite, les dépendances pour lesquelles le gouverneur appartient au segment sont préfixées par OUT\_, celles pour lesquelles le dépendant appartient au segment sont préfixées par IN\_ ; les dépendances internes au segment ne sont pas considérées. Pour illustration, les vecteurs de contexte extraits pour les bisegments encadrés de la figure 1 sont donnés en figure 2<sup>3</sup>. Le dernier élément des vecteurs correspond à la classe de l'exemple pour le classifieur.

Ligne	Phrases alignées
948	finally , aid can enable restructuring , offer training , <b>save</b> jobs and thus know-how .
	finalément , les aides peuvent permettre des restructurations , offrir une formation , <b>sauver</b> des emplois et donc du savoir-faire .
1254	not to <b>save</b> jobs .
	ce n' est pas pour <b>préserver</b> les emplois .
23997	above all we want to <b>save</b> fossil fuels .
	ce que nous voulons surtout , c' est <b>économiser</b> les combustibles fossiles .

FIG. 1 – Exemples de phrases alignées anglais-français issues du corpus Europarl.

Traits lexicaux			Traits morphosyntaxiques			Traits basés sur les dépendances			Traduction
mots	mot-1	mot+1	cat	cat-1	cat+1	OUT_dobj	IN_ccomp	...	
save	,	jobs	V	,	NNS	jobs	enable	-	sauver
save	to	jobs	V	TO	NNS	jobs	NIL	-	préserver
save	to	fossil	V	TO	JJ	fuels	NIL	-	économiser

FIG. 2 – Exemples de vecteurs de contextes extraits pour les exemples de la figure 1.

### 3.3 Intégration dans un système PBSMT

Puisque chaque segment source apparaît dans un contexte qui lui est propre, des entrées uniques dans la table de traduction doivent lui être associées. Une étape de prétraitement construit donc dynamiquement un fichier source sérialisé où chaque mot est rendu unique (ex: belong@37

<sup>2</sup>On peut par exemple s'attendre dans un bon corpus à avoir  $P(\text{sauver}|\text{save}, \{\text{obj} = \text{bank}\}) \gg P(\text{économiser}|\text{save}, \{\text{obj} = \text{bank}\})$

<sup>3</sup>La taille du contexte immédiat a été limitée à un mot, et les cibles des dépendances correspondent aux mots.

t o@38) et une table de traduction associant aux segments impliquant des mots uniques leurs scores standard, leur score contextuel (obtenu en classifiant chaque segment source unique et en normalisant les scores obtenus) et le score binaire caractérisant la traduction la plus probable selon le modèle contextuel. L'extraction des vecteurs contextuels se fait à partir des alignements entre segments et des sorties de l'analyse linguistique visée.

## 4 Expériences en traduction anglais → français

### 4.1 Cadre expérimental

Les expériences rapportées par Stroppa *et al.* (2007) ont été menées avec le classifieur IGTre du package TiMBL (Daelemans *et al.*, 2007). Les gains d'information<sup>4</sup> sont utilisés pour sélectionner d'abord les traits les plus discriminants. Un vote majoritaire est effectué pour déterminer la meilleure classe (i.e. la meilleure traduction du segment source), et il est possible d'obtenir un ensemble pondéré de classes, ce qui après normalisation permet d'obtenir les distributions  $P(e_k|f_k, C(f_k, f))$  recherchées. Les sorties de IGTre étant très sensibles aux faibles différences de valeurs d'informativité entre traits (ce qui est particulièrement le cas pour les traits provenant de dépendances syntaxiques, cf. figure 3), nous avons utilisé le classifieur hybride Tribl de TiMBL. Celui-ci réalise un parcours de l'arbre de décision IGTre pour les traits les plus discriminants, puis utilise une classification de type  $k$ -NN pour les traits moins discriminants.

Nous avons utilisé les débats parlementaires européens du corpus Europarl (Koehn, 2005), et avons limité la taille du corpus d'apprentissage à 100 000 lignes afin de rendre possible l'apprentissage des classifieurs sur les ordinateurs à disposition.<sup>5</sup> Le parseur probabiliste de Stanford (de Marneffe *et al.*, 2006) a été utilisé pour analyser en dépendances la partie anglaise du corpus d'apprentissage et les textes à traduire.

Afin d'obtenir des vecteurs d'apprentissage de taille constante, un nombre fixe de dépendances syntaxiques a été conservé. Nous avons opéré une sélection des 16 dépendances directionnelles (i.e. le gouverneur étant à l'intérieur ou à l'extérieur du segment) menant au gain d'information le plus élevé sur les vecteurs de contexte issus de notre corpus d'apprentissage. La figure 3 décrit les traits retenus, ordonnés par gain d'information décroissant. Nous rapportons également les valeurs obtenues pour les systèmes italien→anglais et chinois→anglais de (Stroppa *et al.*, 2007) pour les traits communs aux deux études.

Comme attendu, les mots d'un segment et ses catégories sont les traits les plus discriminants pour le choix de sa traduction. Les traits lexicaux du contexte immédiat puis les traits morpho-syntaxiques sont les suivants, le contexte droit apparaissant comme plus discriminant que le contexte gauche. Les traits basés sur les dépendances apparaissent moins discriminants individuellement, mais cela est en particulier dû au fait qu'ils ne concernent que des sous-ensembles de segments sources. La dépendance la plus utile est IN\_det, qui permet de choisir en particulier la traduction d'un déterminant, non marqué en genre en anglais, en fonction du nom auquel il se rattache, et dont la traduction en français sera elle marquée en genre. Les dépendances suivantes sont IN\_nsubj et OUT\_nsubj, qui correspondent au cas où il faut choisir la traduction d'un su-

---

<sup>4</sup>Le gain d'information est défini comme la différence d'entropie entre des situations sans connaissance des valeurs d'un trait et avec connaissance de ses valeurs, et renseigne donc sur le caractère discriminant des traits.

<sup>5</sup>L'approche suivie produit des classifieurs construits sur la base de nombreux exemples, ainsi que des tables de traduction volumineuses du fait du caractère unique des segments sources qu'elles contiennent.

jet en fonction de son verbe (le type de corpus utilisé contient beaucoup de formes verbales à la 3ème personne du présent qui sont marquées par une flexion) et d'un verbe et fonction de son sujet, ce qui reflète des accords grammaticaux importants en français. Des dépendances telles que IN\_amod et IN\_dobj, qui correspondent à la traduction d'un adjectif connaissant le nom qu'il modifie et à la traduction d'un verbe connaissant son objet direct, concernent elles davantage des désambiguïisations de sens.

Traits	Gain d'information			Description
	en→fr	it→en	zh→en	
PHR	7,29	7,82	6,74	mots du segments
POS	4,08	4,59	3,23	catégories morphosyntaxiques des mots du segment
W+1	3,72	4,24	3,73	premier mot sur la droite du segment
W-1	3,45	4,09	3,21	premier mot sur la gauche du segment
W+2	2,98	3,19	2,90	second mot sur la droite du segment
W-2	2,78	2,84	2,25	second mot sur la gauche du segment
POS+1	1,68	1,75	1,03	catégorie du premier mot sur la droite du segment
POS-1	1,44	1,61	1,18	catégorie du premier mot sur la gauche du segment
POS+2	1,04	0,90	0,75	catégorie du second mot sur la droite du segment
POS-2	0,89	0,94	0,77	catégorie du second mot sur la gauche du segment
IN_det	0,75	-	-	déterminant ; le nom modifié est en dehors du segment ; ex: <i>[the] session</i>
IN_nsubj	0,65	-	-	sujet nominal dont le verbe est en dehors du segment ; ex: <i>[I] declare</i>
OUT_nsubj	0,43	-	-	inverse de IN_nsubj ; ex: <i>I [declare]</i>
IN_amod	0,42	-	-	adjectif ; le nom modifié est en dehors du segment ; ex: <i>[happy] new year</i>
IN_advmod	0,39	-	-	adverbe ; le mot modifié est en dehors du segment ; ex: <i>[truly] dreadful</i>
OUT_prep_of	0,35	-	-	nom modifié par un nom en dehors du segment via la proposition <i>of</i> ; ex: <i>[a number] of countries</i>
OUT_dobj	0,34	-	-	verbe transitif dont l'objet direct est en dehors du segment ; ex: <i>[you requested] a debate</i>
OUT_amod	0,34	-	-	inverse de IN_amod ; ex: <i>happy new [year]</i>
IN_dobj	0,33	-	-	inverse de OUT_dobj ; ex: <i>you requested [a debate]</i>
OUT_det	0,31	-	-	inverse de IN_det ; ex: <i>the [session]</i>
IN_nn	0,28	-	-	nom modifiant un autre nom en dehors du segment ; ex: <i>[Madam] President</i>
IN_prep_of	0,26	-	-	inverse de OUT_prep_of ; ex: <i>a number of [countries]</i>
OUT_nn	0,26	-	-	inverse de IN_nn ; ex: <i>Madam [President]</i>
OUT_advmod	0,21	-	-	inverse de IN_advmod ; ex: <i>truly [dreadful]</i>

FIG. 3 – Description des traits par gain d'information décroissant pour le système anglais → français. Les gains rapportés par Stroppa *et al.* (2007) pour leurs systèmes italien → anglais et chinois → anglais sont indiqués à titre de comparaison.

Le corpus d'apprentissage a été filtré pour ne conserver que les phrases de moins de 60 mots et pour lesquelles une analyse par l'analyseur de Stanford a été possible, soit 95 734 lignes, et des corpus de développement et de test (resp. 475 et 472 lignes) ont subi le même traitement,

à la différence que les vecteurs de contexte ne renseignent pas sur la traduction du segment source. Les alignements ont été obtenus à l'aide de Giza++ et de l'heuristique grow-diag-final-and (Koehn *et al.*, 2003) en limitant la taille des segments à 7 mots. Pour le corpus d'apprentissage, 11,5M vecteurs de contexte ont été obtenus. Les mots des segments étant de loin le trait le plus discriminant, un filtrage basé sur les segments apparaissant effectivement dans les corpus de développement et de test a permis de limiter ce nombre à 3,7M vecteurs.

Trois systèmes de traduction basés sur les segments ont été construits pour la paire anglais → français pour le décodeur PBSMT Moses<sup>6</sup>, en utilisant des modèles de langues trigrammes pour le français appris sur la partie cible du bitexte d'apprentissage<sup>7</sup>. Les poids de la combinaison de modèles ont été ajustés dans tous les cas par l'approche MERT (Och, 2003). Le système BAS correspond à un système baseline Moses. Le système S1, construit sur BAS, reprend les traits de (Stroppa *et al.*, 2007), et S2 ajoute les traits basés sur les dépendances sélectionnées. Afin de permettre un temps d'apprentissage raisonnable, un filtrage ne conservant que les bisegments pour lesquels  $p(e_k|f_k) > 2.10^{-4}$  a été effectué, Gimpel & Smith (2008) rapportant que cela ne dégrade pas les performances de leurs systèmes.

## 4.2 Évaluation

Les résultats d'évaluation obtenus avec les métriques BLEU, NIST et TER<sup>8</sup>, ainsi que les résultats d'une évaluation manuelle, sont présentés en figure 4. Les métriques automatiques indiquent que le système S1 semble légèrement inférieur aux deux autres systèmes, ce qui semble au moins indiquer un rôle positif joué par les traits propres à S2. BAS et S2 ont des performances comparables sur toutes les métriques, les différences observées n'étant pas significatives.

Plusieurs travaux ont critiqué l'adéquation des métriques du type de BLEU à rendre compte de différences fines liées à la prise en compte du contexte (ex. (Giménez & Màrquez, 2007)). Nous avons donc mené une évaluation manuelle sur 100 phrases du corpus de test tirées au hasard<sup>9</sup>. Quatre juges francophones ont classé les sorties des trois systèmes présentées dans un ordre aléatoire, les égalités étant permises. La moyenne des rangs moyens pour l'ensemble des juges semble indiquer une préférence plus marquée pour les systèmes contextuels, avec une performance légèrement supérieure pour S2. Deux juges ont donné des rangs équivalents aux deux systèmes contextuels, les deux autres préférant S2.

	Scores automatiques			Scores manuels			
	BLEU	NIST	TER	rang moyen	% 1er	% 2nd	% 3ème
BAS	<b>31,13</b>	6,73	56.45	1,54	64,66	9,33	26,00
S1	30,23	6,70	56.81	1,39	74,00	12,00	14,00
S2	30,94	<b>6.76</b>	<b>56.33</b>	<b>1,34</b>	<b>74,66</b>	13,33	12,00

FIG. 4 – Résultats des évaluations automatiques et manuelles.

<sup>6</sup><http://www.statmt.org/moses/>

<sup>7</sup>Nos prochaines expériences utiliseront des modèles de langue cible appris sur des volumes de données beaucoup plus importants.

<sup>8</sup>Une seule traduction de référence est disponible par phrase de test.

<sup>9</sup>Pour ces phrases, les trois systèmes ont 8 traductions identiques, BAS et S1 en ont 11, BAS et S2 en ont 22, et S1 et S2 en ont 11.



<b>Src</b>	<i>I think specifically that in going too quickly to six then to fifteen we have displayed our inability to build a Europe which respects the disparities between its initial members.</i>
<b>Réf</b>	<i>je crois précisément qu'en allant déjà trop vite à six , puis à quinze , nous avons démontré notre incapacité à construire une Europe qui respecte les différences de ses premiers membres.</i>
<b>Bas</b>	<i>je pense que, précisément à aller trop vite à six, quinze nous ont fait preuve notre incapacité à construire une Europe qui respecte les différences entre ses membres initiale.</i>
<b>S1</b>	<i>je pense notamment qu'à aller trop vite pour six ensuite à quinze, nous avons hissé notre incapacité à construire une Europe qui respecte les disparités entre ses membres initiaux.</i>
<b>S2</b>	<i>je crois que précisément à aller trop rapidement de six à quinze, nous avons démontré notre incapacité à construire une Europe qui respecte les différences entre ses membres initiaux.</i>
<b>Src</b>	<i>it is our job to continue to support Latvia with the integration of the Russian population.</i>
<b>Réf</b>	<i>il est de notre devoir de continuer d'appuyer la Lettonie sur la question de l'intégration de la population russe.</i>
<b>Bas</b>	<i>c'est notre tâche consiste à continuer à soutenir la Lettonie à l'intégration de la population russe.</i>
<b>S1</b>	<i>c'est notre tâche consiste à continuer à soutenir la Lettonie à l'intégration de la population russe.</i>
<b>S2</b>	<i>notre tâche consiste à continuer à soutenir la Lettonie avec l'intégration de la population russe.</i>

FIG. 5 – Deux traductions produites par nos trois systèmes (**Src** indique la phrase traduite, et **Réf** une traduction de référence possible).

Il est intéressant de noter que nos systèmes contextuels ont tendance à préférer des segmentations impliquant des segments plus longs, ce qui rejoint les observations de Carpuat & Wu (2008). On peut conjecturer que certains scores de traduction pour des segments longs peu fréquents sont renforcés par la présence d'indices contextuels discriminants.

## 5 Discussion et perspectives

Les gains que nous avons obtenus par la prise en compte de dépendances syntaxiques pour la traduction de segments sont modestes au niveau des évaluations automatiques, mais ils correspondent à notre connaissance à la première amélioration en exploitation du contexte d'une phrase source en traduction vers une langue fortement fléchie. En outre, notre évaluation manuelle conforte l'évaluation positive de notre apport, relativement à un système PBSMT standard, et relativement à un système contextuel n'exploitant pas de dépendances syntaxiques. L'inspection des sorties des différents systèmes permet en effet de constater que notre système ne dégrade que très rarement les traductions, et qu'il améliore parfois de façon très notable les traductions comme dans les deux exemples de la figure 5.

Plusieurs raisons peuvent être avancées pour expliquer pourquoi les métriques automatiques ne montrent pas de gains plus marqués. En particulier, les améliorations sensibles concernent peut-être un nombre trop limité de phrases, et les modifications positives effectuées ne correspondent peut-être pas aux formulations contenues dans la traduction de référence utilisée. Par ailleurs, l'analyse en dépendances sur laquelle se base en grande partie notre approche est un traitement sujet à erreur.

Nos travaux futurs vont tout d'abord porter sur les limitations du présent travail. En premier lieu, nous devons travailler sur la résolution de plusieurs problèmes d'implémentation, qui nous ont

par exemple contraints à limiter le nombre d'exemples d'apprentissage pour le classifieur ou à filtrer les entrées de la table de traduction contextuelle. Nous essaierons ensuite d'intégrer une meilleure prise en compte du contexte, par exemple en construisant des classifieurs adaptés aux types de segments (par exemple, segments contenant un verbe transitif et ne contenant pas son objet direct), et nous appliquerons notre technique à plusieurs paires de langues pour pouvoir réaliser des études contrastives. Il sera notamment particulièrement intéressant de comparer le pouvoir discriminant de dépendances syntaxiques comparables en fonction de la paire de langues considérée.

## Références

- ALLAUZEN A., CREGO J., MAX A. & YVON F. (2009). LIMSI's statistical translation systems for WMT'09. In *Proceedings of WMT'09*, Athens, Greece.
- CARPUAT M. & WU D. (2005). Word sense disambiguation vs statistical machine translation. In *Proceedings of ACL*, Ann Arbor, USA.
- CARPUAT M. & WU D. (2007). Context-dependent phrasal translation lexicons for statistical machine translation. In *Proceedings of Machine Translation Summit XI*, Copenhagen, Denmark.
- CARPUAT M. & WU D. (2008). Evaluation of context-dependent phrasal translation lexicons for statistical machine translation. In *Proceedings of LREC 2008*, Marrakech, Morocco.
- DAELEMANS W., ZAVREL J., VAN DER SLOOT K. & VAN DEN BOSCH A. (2007). *TIMBL: Tilburg Memory Based Learner, version 6.1, Reference Guide*. Rapport interne, ILK 07-xx.
- DE MARNEFFE M.-C., MACCARTNEY B. & MANNING C. D. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC'06*, Genoa, Italy.
- GIMÉNEZ J. & MÀRQUEZ L. (2007). Context-aware discriminative phrase selection for statistical machine translation. In *Proceedings of WMT at ACL*, Prague, Czech Republic.
- GIMPEL K. & SMITH N. A. (2008). Rich source-side context for statistical machine translation. In *Proceedings of WMT at ACL*, Columbus, USA.
- KOEHN P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of MT Summit*, Phuket, Thailand.
- KOEHN P., OCH F. J., & MARCU D. (2003). Statistical phrase-based translation. In *Proceedings of NAACL/HLT*, Edmonton, Canada.
- LANGLAIS P., GOTTI F. & PATRY A. (2006). De la chambre des communes à la chambre d'isolement: adaptabilité d'un système de traduction basé sur les segments. In *Actes de TALN*, Louvain, Belgique.
- LAVECCHIA C., SMAILI K., LANGLOIS D. & HATON J.-P. (2008). Une alternative aux modèles de traduction statistique d'IBM: Les triggers inter-langues. In *Actes de TALN*, Avignon, France.
- OCH F. J. (2003). Minimum error rate training for statistical machine translation. In *Proceedings of ACL*, Sapporo, Japan.
- STROPPA N., VAN DEN BOSCH A. & WAY A. (2007). Exploiting source similarity for SMT using context-informed features. In *Proceedings of TMI*, Skövde, Sweden.
- VILAR D., XU J., D'HARO L. F. & NEY H. (2006). Error Analysis of Statistical Machine Translation Output. In *Proceedings of LREC*, Genoa, Italy.