

Génération de reformulations locales par pivot pour l'aide à la révision

Aurélien Max
LIMSI-CNRS et Université Paris-Sud 11
Orsay, France
aurelien.max@limsi.fr

Résumé. Cet article présente une approche pour obtenir des paraphrases pour de courts segments de texte qui peuvent aider un rédacteur à reformuler localement des textes. La ressource principale utilisée est une table d'alignements bilingues de segments d'un système de traduction automatique statistique. Un segment marqué par le rédacteur est tout d'abord traduit dans une langue pivot avant d'être traduit à nouveau dans la langue d'origine, ce qui est permis par la nature même de la ressource bilingue utilisée sans avoir recours à un processus de traduction complet. Le cadre proposé permet l'intégration et la combinaison de différents modèles d'estimation de la qualité des paraphrases. Des modèles linguistiques tentant de prendre en compte des caractéristiques des paraphrases de courts segments de textes sont proposés, et une évaluation est décrite et ses résultats analysés. Les domaines d'application possibles incluent, outre l'aide à la reformulation, le résumé et la réécriture des textes pour répondre à des conventions ou à des préférences stylistiques. L'approche est critiquée et des perspectives d'amélioration sont proposées.

Abstract. In this article, we present a method to obtain paraphrases for short text spans that can be useful to help a writer in reformulating text. The main resource used is a bilingual phrase table containing aligned phrases, a common resource in statistical machine translation. The writer can mark a segment for paraphrasing, and this segment is first translated into a pivot language before being back-translated into the original language, which is possible without performing a full translation of the input. Our proposed framework allows integrating and combining various models for estimating paraphrase quality. We propose linguistic models which permits to conduct empirical experiments about the characteristics of paraphrases for short text spans. Application domains include, in addition to paraphrasing aids, summarization and rephrasing of text for conforming to conventional or stylistic guidelines. We finally discuss the limitations of our work and describe possible ways of improvement.

Mots-clés : Paraphrase, Traduction Automatique Statistique basée sur les segments, Aide à la rédaction.

Keywords: Paraphrasing, Phrase-Based Statistical Machine Translation (PBSMT), Authoring aids.

1 Introduction

La vie d'un texte est souvent pleine de rebondissements. Un auteur peut le réviser pour lui apporter des corrections, le rendre plus court, l'adapter à un lectorat voire le traduire dans une autre langue. Lors de la phase de maturation d'un texte, celui-ci peut subir de nombreux changements avant de prendre sa forme définitive. À contenu informationnel (plus ou moins) constant, l'auteur cherche à tendre vers des formulations qui seront plus adaptées au contexte en termes de correction linguistique et de compréhensibilité. La question de la génération automatique de reformulations, même locales, et de l'évaluation de leur adaptation à leur contexte semble donc particulièrement pertinente alors que les aides à la rédaction sont encore trop peu sensibles à la nature linguistique des textes manipulés.

Des travaux en linguistique distinguent la reformulation à *visée explicative*, qui tente d'explicitement le sens, et la reformulation à *visée imitative*, qui tente de substituer des formulations équivalentes (Fuchs, 1994). Ce dernier cas pose la difficile question de la prise en compte d'un *seuil de distortion* au-delà duquel une reformulation n'est plus admissible. Les systèmes de transformation automatique de texte butent en effet sur la prise en compte d'un tel seuil, qui n'est pas explicité dans les principales mesures d'évaluation utilisées, telles que BLEU (Papineni *et al.*, 2002) pour la traduction automatique ou ROUGE (Lin, 2004) pour le résumé. Une certaine modélisation de la reformulation mène à des techniques d'analyse et de reconnaissance de textes qui sont des paraphrases les uns des autres à des fins d'extraction d'informations, mais où cette notion de seuil est souvent éclipsée au profit du rappel.

Dans cet article, nous présentons une approche permettant de proposer à un rédacteur des reformulations pour de courts segments de texte (par ex. *dans de bonnes conditions, maintenir le contact, la plus étroite*) au cours de l'activité de révision. Ce type d'aide reprend les principes de certains travaux visant à faire des suggestions lexicales, par exemple en enrichissant des dictionnaires avec des index d'associations, telles que des mots liés thématiquement (Ferret & Zock, 2006). Dans cet article, nous commençons par décrire brièvement les travaux existant en paraphrasage automatique, puis nous décrivons les fondements de notre approche inspirée de (Bannard & Callison-Burch, 2005), qui consiste à utiliser des traductions de segments en pivot pour produire des reformulations et sélectionner parmi celles-ci celles qui sont préférées par différents types de modèles. Une évaluation est décrite et ses résultats sont analysés, ce qui met en évidence le caractère prometteur de l'approche proposée et suggère de nouvelles voies de recherche. En particulier, nous soulignons l'importance d'utiliser le contexte de la phrase d'origine contenant le segment à paraphraser pour le choix des segments pivots, importance qu'on retrouve dans le domaine de la traduction automatique.

2 Travaux antérieurs

La possibilité d'utiliser de grands corpus comparables et parallèles a suscité récemment une vague de travaux autour du paraphrasage. Des travaux ont par exemple porté sur l'apprentissage et l'utilisation de paraphrases pour améliorer des tâches telles que la recherche d'informations précises (Duclaye *et al.*, 2003) ou la déduction sur textes, et des applications texte-à-texte, telle que la traduction automatique (Callison-Burch, 2007), ou encore l'évaluation de ce type d'applications en traduction (Kauchak & Barzilay, 2006) ou en résumé (Zhou *et al.*, 2006).

Les systèmes de génération automatique de texte non-déterministes offrent un cadre naturel pour la génération de paraphrases, dont les plus « naturelles » peuvent être sélectionnées par un modèle de langue (Landkilde & Knight, 1998). Si un point fort de ce type d'approche est qu'il est possible de produire des paraphrases de nature très différentes, la construction d'un tel système permettant de produire une grande variabilité de textes en sortie est difficile, et ces systèmes attendent des entrées qui sont soit sous forme de concepts, soit sous forme de spécifications de phrases.

Des alignements appris à partir de corpus comparables (monolingues) permettent de générer des reformulations en se passant d'un tel générateur. Par exemple, (Barzilay & Lee, 2003) apprennent des treillis factorisant l'ensemble des reformulations possibles au niveau des phrases, ce qui est rendu possible par la nature même de leur corpus d'apprentissage (documents journalistiques décrivant les mêmes événements). Pour une phrase en entrée, si un treillis qui lui correspond est trouvé, celui-ci est utilisé pour produire un ensemble de formulations correspondant aux chemins du treillis. Si ce type d'approche peut parfois produire de très bonnes reformulations phrastiques, une bonne couverture des phénomènes linguistiques s'avère difficile à obtenir. En outre, des reformulations plus locales sont moins susceptibles d'altérer le sens des phrases. (Quirk *et al.*, 2004) proposent une approche consistant à apprendre un système de traduction statistique sur un corpus monolingue de phrases alignées automatiquement à partir d'un corpus comparable, qui fonctionne par décodage monotone et donc opère par reformulations locales. Si cette absence de transformations profondes est souvent perçue comme l'une des principales limitations des systèmes de traduction statistiques actuels, elle permet néanmoins de limiter les risques de déformation du sens. D'autres travaux ont exploité des corpus parallèles monolingues, comme différentes traductions d'un même texte, pour apprendre des treillis permettant de générer des paraphrases (Pang *et al.*, 2003).

Si ces approches s'attaquent toutes au paraphrasage d'énoncés complets, il est également intéressant de proposer des reformulations pour des unités plus petites. À partir de corpus parallèles bilingues, ressources traditionnelles pour l'apprentissage des systèmes de traduction statistique, (Bannard & Callison-Burch, 2005) ont proposé une approche pour le paraphrasage de courts segments de mots par traduction dans une langue puis par rétro-traduction et sélection dans la langue d'origine. Le problème de la validité d'une paraphrase en contexte pose notamment la question de la prise en compte de son contexte d'apparition, problème qui a récemment été abordé en traduction automatique statistique (Stroppa *et al.*, 2007). Des travaux existent au niveau des paraphrases lexicales, comme ceux de (Connor & Roth, 2007) qui proposent un classifieur pour décider si un mot peut en remplacer un autre dans un certain contexte.

3 Reformulation de courts segments de texte par pivot

3.1 Description générale de l'approche

La traduction automatique statistique (*Statistical Machine Translation* (SMT)) est fondée sur l'apprentissage et l'utilisation d'alignements entre mots appris sur des corpus parallèles bilingues. La SMT basée sur des segments (Koehn *et al.*, 2003) a étendu la notion d'alignements entre segments de taille quelconque, ce qui permet une meilleure prise en compte du contexte local, sans que ces segments n'aient à correspondre à une segmentation motivée linguistiquement. Des techniques d'alignement permettent d'apprendre des probabilités de traduction de

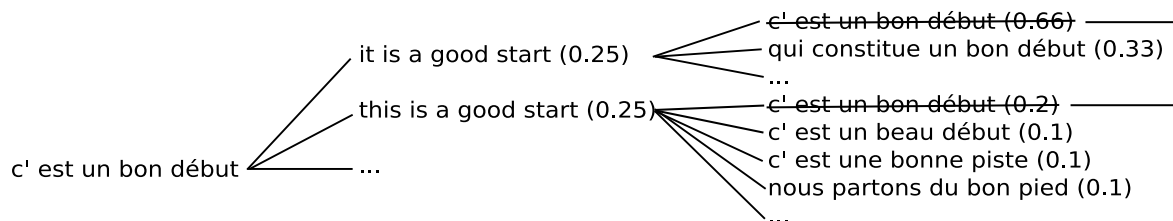


FIG. 1 – Exemple de paraphrasage de segment par consultation de tables de traduction. Les valeurs indiquent la probabilité conditionnelle du segment étant donné le segment à sa gauche.

segments. Par exemple, (Koehn *et al.*, 2003) partent d’alignements $1 \rightarrow N$ entre mots pour une paire de langues dans les deux directions, symétrisent ces alignements en prenant leur intersection, puis construisent des alignements $N \rightarrow M$ entre mots de façon incrémentale en ajoutant des éléments de l’union des deux alignements initiaux. Pour qu’un alignement soit construit à partir des alignements existant à une itération précédente, tous les mots d’un segment source doivent être uniquement alignés avec les mots d’un segment cible et inversement. À partir de l’ensemble des segments alignés extraits d’un corpus parallèle, les probabilités de traduction sont en général estimées à partir de fréquences relatives correspondant au rapport du nombre de fois où un segment source s et un segment cible c sont alignés sur le nombre d’occurrences du segment source : $P(c|s) = \frac{\text{compte}(c,s)}{\text{compte}(s)}$.

Des tables de traduction de segments sont créées en utilisant notamment ces probabilités entre segments. (Bannard & Callison-Burch, 2005) ont proposé d’utiliser ce type de probabilités à des fins de paraphrasage, en définissant une *probabilité de paraphrasage* entre des segments seg_1 et seg_2 qui utilise l’alignement qu’entretiennent ces deux segments avec un segment pivot dans une autre langue, *pivot* :

$$\hat{seg}_2 = \arg \max_{seg_2 \neq seg_1} P(seg_2|seg_1) = \arg \max_{seg_2 \neq seg_1} \sum_{pivot} P(seg_2|pivot)P(pivot|seg_1) \quad (1)$$

Dans l’équation 1, la recherche se fait en excluant le segment d’origine seg_1 et en sommant sur l’ensemble des pivots possibles pour une paire (seg_1, seg_2) . L’exemple de la figure 1 illustre le processus de construction de paraphrases d’un segment par pivot.

Outre la probabilité de paraphrasage décrite ci-dessus, (Callison-Burch, 2007) a utilisé différentes sources d’information pour mesurer l’importance de différents facteurs dans la qualité des reformulations¹ :

- qualité des alignements entre segments : la construction manuelle d’un corpus dont les alignements sont faits par un humain permet une amélioration significative de la qualité des paraphrases obtenues.
- utilisation de plusieurs corpus alignés : l’utilisation de plusieurs paires de langues pour obtenir les paraphrases (les pivots sont donc multilingues) permet d’éliminer partiellement les défauts d’alignements et d’obtenir une meilleure performance.
- contrôle du sens des segments : sans qu’il ne s’agisse de désambiguïsation sémantique des

¹Dans ses expériences, Callison-Burch utilise 46 segments en anglais choisis aléatoirement parmi des expressions composées de plusieurs mots de WordNet (ex : *at work*, *concentrate on*, *big business*), et 289 phrases qui les contiennent. L’évaluation est alors effectuée par deux juges qui doivent attribuer aux reformulations des scores de conservation du sens (*adequacy*) et de caractère naturel (*fluency*).

segments (qui peuvent avoir, comme les mots, plusieurs sens dépendant de leur contexte d'apparition), paraphraser des segments appartenant à des phrases alignées avec des phrases dans une autre langue permet de restreindre les traductions possibles du segment, et améliore sensiblement les résultats.

- contrôle de la vraisemblance des paraphrases produites : l'utilisation d'un modèle de langue trigramme permet d'améliorer la grammaticalité des paraphrases produites, mais a un impact négatif sur la conservation du sens.

Notre travail se base sur cette approche en proposant une intégration de différents modèles pour aider à la sélection des paraphrases. L'utilisation d'un corpus de paraphrases évaluées par un humain peut alors servir à étudier la contribution de chacun des modèles, en particulier dans une tâche d'aide à la reformulation. Le score d'une reformulation peut être évalué en utilisant une combinaison linéaire de logarithmes correspondant à des scores de modèles telle que traditionnellement utilisée en SMT² (M représente les modèles utilisés, et λ_m et h_m représentent respectivement le poids et la fonction de score d'un modèle) :

$$s\hat{e}g_2 = \arg \max_{seg_2} P(seg_2|seg_1) = \arg \max_{seg_2} \sum_{m \in M} \lambda_m h_m(seg_1, seg_2) \quad (2)$$

3.2 Modèles pour la sélection des reformulations

Nous présentons dans cette section les modèles que nous avons utilisés pour évaluer les reformulations.³

Modèle pivot (PIV) Nous utilisons un score dérivé de la probabilité de paraphrasage de (Barnard & Callison-Burch, 2005) en utilisant une seule langue pivot, qui nous servira de référence (*baseline*) :

$$Piv(p_1, p_2) = \log\left(\sum_{pivot} P(p_2|pivot)P(pivot|p_1)\right) \quad (3)$$

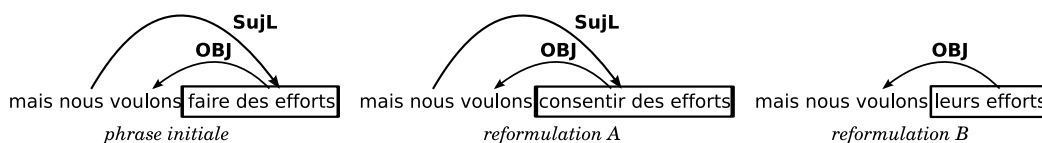
Modèle de conservation de dépendances (DEP) Ce modèle vise à prendre en compte une certaine proportion des relations de dépendance qui sont conservées par la reformulation. Il prend en compte les dépendances externes au segment avant et après reformulation (resp. \mathcal{D}_1^{extra} et \mathcal{D}_2^{extra}) sous la forme (*dépendance, gouverneur, dépendant*), ainsi que les dépendances entrant ou sortant du segment avant et après reformulation (resp. \mathcal{D}_1^{inter} et \mathcal{D}_2^{inter}) sous la forme (*dépendance, cible_hors_segment*) :

$$Dep(p_1, p_2, context) = \log\left(\frac{1 + |\mathcal{D}_1^{extra} \cap \mathcal{D}_2^{extra}| + |\mathcal{D}_1^{inter} \cap \mathcal{D}_2^{inter}|}{1 + |\mathcal{D}_1^{extra}| + |\mathcal{D}_1^{inter}|}\right) \quad (4)$$

²L'approche choisie pourrait suggérer l'utilisation de systèmes de SMT complets pour le paraphrasage par pivot d'énoncés entiers, mais les performances des systèmes actuels, qui butent notamment sur la prise en compte du contexte source et sur la conservation du sens et de la grammaticalité en cible, rendent pour le moment cette approche problématique. (Quirk *et al.*, 2004) attaquent eux ce problème avec un système SMT appris avec des alignements monolingues, ce qui constitue une voie de recherche intéressante.

³Nous avons également fait des expériences avec des modèles basés sur les catégories morphosyntaxiques des mots, comme par exemple modélisant la probabilité de la suite de catégories d'un segment étant donné son contexte constitué des catégories à gauche et à droite du segment, mais ces modèles n'ont permis d'améliorer les résultats que nous présenterons en section 4.

Le schéma ci-dessous illustre les cas où deux dépendances entrante et sortante sont retrouvées (*reformulation A*) et où une dépendance entrante n'est pas retrouvée (*reformulation B*).



Modèle de lemmes (LEM) Ce modèle mesure une certaine proportion de lemmes de mots pleins communs au segment initial p_1 et au segment après reformulation p_2 (resp. \mathcal{L}_1 et \mathcal{L}_2) :

$$Lem(p_1, p_2) = \log \left(\frac{1 + |\mathcal{L}_2| - |\mathcal{L}_1 \cap \mathcal{L}_2|}{1 + |\mathcal{L}_1 \cup \mathcal{L}_2|} \right) \quad (5)$$

4 Expériences et évaluation

Les tables de traduction utilisées ont été apprises par la méthode décrite dans la section précédente sur les parties française et anglaise du corpus Europarl⁴, qui comprend environ 1,2 million de phrases alignées. Les tables contiennent environ 42 millions de segments alignés (d'une taille maximale de 10 éléments (*tokens*)) avec des scores de traduction dans les deux directions. Nous avons choisi de travailler avec le français comme langue principale (et donc l'anglais comme seule langue pivot pour le moment), ce qui rend plus manifeste le besoin d'utiliser des segments rendant compte des dépendances ainsi que des accords entre mots qui en découlent. Nous avons utilisé l'analyseur syntaxique robuste pour le français SYNTAX (Bourigault *et al.*, 2005) pour obtenir une segmentation en unités lexicales, la lemmatisation de ces unités et une analyse en dépendances. Outre les modèles décrits en section 3.2, nous avons utilisé un modèle de langue 5-grammes (5GRAM) appris sur la partie française du corpus avec un lissage Kneser-Ney.

Nous avons constitué manuellement un corpus de test de 82 phrases issues d'une partie du corpus Europarl n'ayant pas servi à l'apprentissage des modèles de traduction et du modèle de langue. Pour chacune de ces phrases, un évaluateur à qui il était demandé de simuler la tâche d'un rédacteur en phase de révision a choisi un segment d'intérêt pour la reformulation, celui-ci étant accepté s'il appartenait à la table de traduction français \rightarrow anglais (ex : *dans de bonnes conditions, maintenir le contact, attendent avec impatience, la plus étroite*, etc.).

Nous souhaitons disposer d'un corpus pouvant servir de référence pour différentes expériences impliquant plusieurs modèles. Ne disposant pas d'une métrique automatique acceptable pour évaluer la qualité des reformulations, et souhaitant ne pas avoir à conduire plusieurs évaluations manuelles, nous avons construit automatiquement un premier ensemble de reformulations pour chaque couple (phrase, segment à paraphraser). Afin de limiter le travail d'évaluation, nous avons choisi de ne retenir au plus que les 20 premières reformulations en utilisant PIV seul (ce qui de fait introduit un biais dans nos expériences, en restreignant l'ensemble des reformulations considérées). Pour chacune des 1648 phrases à évaluer, les juges avaient à indiquer les scores suivants sur une échelle de 1 à 5 :

- **score de grammaticalité** : indique si la paraphrase est grammaticalement correcte en l'état (5), nécessite différents niveaux de révision (4 à 2) ou est complètement inexploitable (1).

⁴Version du corpus utilisée pour la tâche WMT08 (<http://www.statmt.org/wmt08>).

	Résultat au premier rang				Résultats au 8 premiers rangs			
	gram.	sens	réduct.	moy.	gram.	sens	réduct.	moy.
PIV (baseline)	4.46	4.18	3.62	4.09	11.66	10.60	9.10	10.45
5GRAM	4.28	3.62	3.45	3.78	11.52	9.72	9.20	10.15
DEP	4.35	3.68	3.43	3.82	11.47	9.70	8.91	10.03
LEM	4.05	3.21	3.28	3.51	10.68	8.74	8.59	9.33
PIV+5GRAM	4.65	4.06	3.82	4.18	11.95	10.30	9.58	10.61
PIV+DEP	4.58	4.27	3.66	4.17	11.85	10.68	9.38	10.64
PIV+LEM	4.37	4.00	3.76	4.05	11.53	10.14	9.56	10.41
5GRAM+DEP	4.49	3.81	3.68	3.99	11.79	9.95	9.51	10.42
5GRAM+LEM	4.28	3.59	3.56	3.81	11.46	9.57	9.37	10.13
PIV+5GRAM+DEP	4.65	4.05	3.92	4.21	12.04	10.35	9.71	10.70
PIV+5GRAM+LEM	4.61	4.02	3.97	4.20	11.86	10.11	9.74	10.57
PIV+DEP+LEM	4.57	4.17	4.02	4.25	11.82	10.33	9.81	10.65
5GRAM+DEP+LEM	4.37	3.69	3.64	3.90	11.65	9.75	9.54	10.31
PIV+5GRAM+DEP+LEM	4.68	4.09	4.05	4.27	12.01	10.24	9.87	10.71

FIG. 2 – Résultats de l'évaluation automatique sur les 82 phrases de notre corpus d'évaluation (les notes au premier rang sont entre 1 et 5, et entre 2.71 et 13.59 pour les 8 premiers rangs)

- **score de conservation du sens** : indique si la paraphrase reprend exactement le sens de la phrase d'origine (5), comporte différents niveaux d'ajout ou de retrait d'information (4 à 2) ou est d'un sens complètement différent (1).
- **score d'intérêt pour un rédacteur** : indique si la paraphrase pourrait être réutilisée en l'état pour une reformulation (5), avec une modification mineure (4), présente des éléments intéressants qui peuvent participer à une reformulation (3), présente des éléments qui peuvent suggérer une reformulation (2) ou correspond à du bruit (1).

Les scores sont obtenus en prenant la moyenne des scores de deux juges francophones.⁵ Le tableau de la figure 2 présente les résultats obtenus par différentes combinaisons de modèles pour notre corpus d'évaluation.⁶ Nous avons ici décidé d'utiliser des poids λ_m uniformes pour la combinaison des modèles, leur optimisation automatique en fonction de caractéristiques choisies par un utilisateur faisant partie de nos travaux futurs. Le deuxième groupe de colonnes indique les moyennes des différents scores et leur moyenne pour la reformulation classée au premier rang. Le troisième groupe de colonnes donne un score combinant les scores des 8 meilleurs résultats⁷, calculé comme suit : $\sum_{rang=1}^8 \frac{1}{rang} score(rang)$.

En ce qui concerne la grammaticalité des paraphrases, les modèles PIV et 5GRAM sont les meilleurs contributeurs. La performance de PIV peut s'expliquer par la nature des alignements, dont les plus forts peuvent correspondre à des patrons morpho-syntaxiques cohérents. La contribution d'un modèle de langue était attendue, mais on constate que 5GRAM seul a une performance légèrement inférieure à DEP. La combinaison PIV+5GRAM mène cependant à une

⁵L'écart-type des différences de scores entre les deux juges est de 0.59 pour la grammaticalité, 0.7 pour la conservation du sens, et 0.8 pour l'intérêt pour un rédacteur, ce qui rend assez bien compte des différents niveaux de subjectivité de ces scores.

⁶Contrairement à ce qui se passe en traduction statistique sur des phrases complètes, la taille de notre espace de recherche permet ici une exploration complète, et l'on a donc la garantie de trouver le meilleur score global (parmi les résultats présélectionnés par PIV).

⁷Ce type de mesure est motivé par la finalité de nos travaux, i.e. proposer un nombre raisonnable de suggestions de reformulation à un rédacteur.

meilleure performance que la combinaison PIV+DEP, ce qui semble indiquer que PIV et DEP opèrent des sélections plus similaires que PIV et LM. À l'examen des données, il apparaît qu'obtenir des reformulations grammaticales requiert parfois des modifications en dehors du segment. Il serait possible d'essayer d'étendre les segments automatiquement en cherchant à obtenir des reformulations de meilleure qualité, mais en conservant l'approche actuelle, on peut remarquer qu'un critère strict de grammaticalité n'est pas nécessairement le plus important pour le rédacteur.

L'évaluation au niveau de la conservation du sens fait ressortir qu'il est difficile de faire mieux que la baseline PIV. Il s'agit cependant certainement là d'un biais imputable au corpus d'apprentissage des tables de traduction (les débats parlementaires européens), qui contient vraisemblablement peu d'ambiguïtés sémantiques au niveau des segments dont les scores d'alignement sont les plus forts. Si la combinaison avec d'autres modèles a plutôt tendance à dégrader les scores, car ils entraînent vraisemblablement une déformation du sens en préférant des énoncés grammaticaux, le modèle DEP entraîne une amélioration lorsque combiné avec PIV, alors que seul il obtient une performance bien moindre. La conservation des relations de dépendances, et en particulier du type de celles-ci pour les dépendances entrant ou sortant dans le segment paraphrasé, participe certainement à cela.⁸ La dégradation observée avec PIV+5GRAM est cohérente avec les résultats obtenus par (Callison-Burch, 2007). Les scores observés en utilisant le modèle LEM semblent indiquer que les reformulations contenant plusieurs mots pleins différents correspondent fréquemment à des sens jugés différents.

Les meilleurs résultats concernant l'intérêt pour le rédacteur sont obtenus en combinant tous les modèles. On constate qu'une mauvaise grammaticalité ou une mauvaise conservation du sens ont tendance à pénaliser ce score, ce qui est cohérent avec les directives d'évaluation. Le modèle 5GRAM se combine à nouveau favorablement avec PIV, et toutes les combinaisons impliquant PIV améliorent le score obtenu par ce modèle seul. Le modèle LEM permet d'améliorer les résultats, vraisemblablement parce qu'il permet de favoriser les reformulations les plus différentes du segment d'origine, qu'un rédacteur peut juger plus utiles qu'un segment très proche. Il permet ainsi parfois de sélectionner des reformulations qui obtiennent par ailleurs de mauvais scores en grammaticalité et/ou en conservation du sens, comme dans l'exemple suivant :

Phrase d'origine : *ceux qui viennent dire ici que l' europe ne doit pas adopter de directive énonçant les valeurs que nous voulons tous défendre se trompent .*

Reformulation : *ceux qui viennent dire ici que l' europe ne doit pas adopter de directive énonçant les valeurs que nous voulons tous défendre devrions à nouveau réfléchir .*

L'évaluation prenant en compte les 8 premiers résultats ne fait pas ressortir de différences significatives avec le résultat au premier rang. Les exemples ci-dessous illustrent deux cas où les premières propositions du système en combinant tous les modèles obtiennent de très bons scores :

Phrase initiale : *et , à cet égard bien précis , je suis d' accord avec vous , monsieur le commissaire*

Reformulations : *je partage votre point de vue, je pense comme vous, je partage votre avis, je vous rejoins*

⁸Ce type de modèle dépend cependant très fortement de la robustesse de l'analyseur syntaxique utilisé pour analyser des énoncés plus ou moins corrects.

Phrase initiale : *notre voisin s' est efforcé résolument pendant la décennie écoulée de remplir les critères de l' adhésion .*

Reformulations : *satisfaire aux exigences, satisfaire aux conditions, répondre aux exigences, respecter les exigences, répondre aux critères, remplir les conditions, satisfaire aux critères*

5 Conclusions et perspectives

Dans cet article, nous avons décrit une approche inspirée de (Bannard & Callison-Burch, 2005) pour produire des reformulations locales en utilisant des tables de traduction de segments et en combinant les scores de différents modèles. Nos évaluations, basées sur trois critères dont un concernant directement l'intérêt des reformulations proposées pour le rédacteur, ont plus particulièrement permis de dégager la contribution de notre modèle basé sur la conservation des relations de dépendance entre un énoncé et ses reformulations. Cependant, l'approche proposée comporte de nombreuses limitations, que nous nous efforcerons de corriger dans nos travaux ultérieurs.

La principale source d'amélioration selon nous consiste à mieux prendre en compte le contexte du segment à paraphraser dans la phrase source. Alors que (Callison-Burch, 2007) a montré l'importance de la prise en compte du contexte pour la génération de paraphrases, ses expériences se sont basées sur un choix de pivot issu directement d'un corpus parallèle bilingue, ce qui revient à faire une désambiguïsation supervisée. Des travaux récents en traduction statistique tels que (Stroppa *et al.*, 2007) ont montré la possibilité d'intégrer dans des systèmes basés sur des segments des modèles permettant d'améliorer la sélection des segments cibles.⁹ Nous travaillons par ailleurs sur ces aspects dans le cadre d'un système de traduction statistique basé sur des segments, et espérons que les résultats bénéficieront au travail présenté ici.

Nous allons par ailleurs poursuivre l'analyse de nos données pour proposer de nouveaux modèles, permettant par exemple de regrouper des sens dans les segments. Dans le cadre d'une intégration dans un système d'aide à la reformulation, nous serons intéressé par la possibilité de corriger les paraphrases produites, pour respecter par exemple des contraintes d'accords grammaticaux. Il nous semble également intéressant de travailler sur le problème du repérage des segments qui seraient de bons candidats à la reformulation (gallicismes en anglais, phraséologie mal adaptée à un genre documentaire, etc.). En outre, nous aurons à considérer l'intégration de segments alignés provenant de corpus comparables, qui sont des ressources beaucoup plus faciles à construire que les corpus alignés. Enfin, ce type de recherche devrait tendre vers plus de généralisation dans le processus de reformulation, et la possibilité de produire des paraphrases plus profondes au niveau d'énoncés, voire de discours. Un type d'application particulièrement intéressant serait alors l'aide à la condensation d'articles scientifiques pour tenir dans des contraintes de taille imposées.

Remerciements L'auteur remercie Daniel Déchelotte pour la préparation de certaines données utilisées dans ce travail, et Didier Bourigault pour l'utilisation de SYNTAX.

⁹Il est possible cependant que la nature des corpus utilisés ne permette pas d'amélioration très sensible des résultats, car le besoin de désambiguïsation est peut-être assez faible dans l'ensemble.

Références

- BANNARD C. & CALLISON-BURCH C. (2005). Paraphrasing with bilingual parallel corpora. In *Proceedings of ACL*, Ann Arbor, USA.
- BARZILAY R. & LEE L. (2003). Learning to paraphrase : an unsupervised approach using multiple-sequence alignment. In *Proceedings of NAACL/HLT*, Edmonton, Canada.
- BOURIGAULT D., FABRE C., FRÉROT C., JACQUES M.-P. & OZDOWSKA S. (2005). Syntex, analyseur syntaxique de corpus. In *Actes de TALN*, Dourdan, France.
- CALLISON-BURCH C. (2007). *Paraphrasing and Translation*. PhD thesis, University of Edinburgh.
- CONNOR M. & ROTH D. (2007). Context sensitive paraphrasing with a single unsupervised classifier. In *Proceedings of ECML*, Warsaw, Poland.
- DUCLAYE F., COLLIN O. & YVON F. (2003). Apprentissage automatique de paraphrases pour l'amélioration d'un système de questions-réponses. In *Actes de TALN*, Batz-sur-mer, France.
- FERRET O. & ZOCK M. (2006). Enhancing electronic dictionaries with an index based on associations. In *Proceedings of COLING/ACL*, Sydney, Australia.
- FUCHS C. (1994). *Paraphrase et énonciation*. Ophrys, Paris.
- KAUCHAK D. & BARZILAY R. (2006). Paraphrasing for automatic evaluation. In *Proceedings of NAACL/HLT*, New York, USA.
- KOEHN P., OCH F. J., & MARCU D. (2003). Statistical phrase-based translation. In *Proceedings of NAACL/HLT*, Edmonton, Canada.
- LANDKILDE I. & KNIGHT K. (1998). Generation that exploits corpus-based statistical knowledge. In *Proceedings of COLING/ACL*, Montréal, Canada.
- LIN C.-Y. (2004). Rouge : a package for automatic evaluation of summaries. In *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, Barcelona, Spain.
- PANG B., KNIGHT K. & MARCU D. (2003). Syntax-based alignment of multiple translations : Extracting paraphrases and generating new sentences. In *Proceedings of NAACL/HLT*, Edmonton, Canada.
- PAPINENI K., ROUKOS S., WARD T. & ZHU W.-J. (2002). Bleu : a method for automatic evaluation of machine translation. In *Proceedings of ACL*, Philadelphia, USA.
- QUIRK C., BROCKETT C. & DOLAN W. B. (2004). Monolingual machine translation for paraphrase generation. In *Proceedings of EMNLP*, Barcelona, Spain.
- STROPPA N., VAN DEN BOSCH A. & WAY A. (2007). Exploiting source similarity for smt using context-informed features. In *Proceedings of TMI*, Skövde, Sweden.
- ZHOU L., LIN C.-Y., MUNTEANU D. S. & HOVY E. (2006). Paraeval : Using paraphrases to evaluate summaries automatically. In *Proceedings of NAACL/HLT*, New York, USA.