

## **Ressources terminologiques et traduction probabiliste: premiers pas positifs vers un système adaptatif**

Philippe Langlais  
RALI/DIRO - Université de Montréal  
C.P. 6128, succursale Centre-ville  
Montréal (Québec)  
Canada, H3C 3J7  
felipe@iro.umontreal.ca

### **Mots-clefs – Keywords**

Traduction statistique, adaptabilité, terminologie  
Statistical machine translation, adaptativity, terminology

### **Résumé - Abstract**

Cette dernière décennie a été le témoin d'importantes avancées dans le domaine de la traduction statistique (TS). Aucune évaluation fine n'a cependant été proposée pour mesurer l'adéquation de l'approche statistique dans un contexte applicatif réel.

Dans cette étude, nous étudions le comportement d'un engin de traduction probabiliste lorsqu'il traduit un texte de nature très éloignée de celle du corpus utilisé lors de l'entraînement. Nous quantifions en particulier la baisse de performance du système et développons l'idée que l'intégration de ressources terminologiques dans le processus est une solution naturelle et salutaire à la traduction. Nous décrivons cette intégration et évaluons son potentiel.

The past decade witnessed exciting work in the field of Statistical Machine Translation (SMT). However, accurate evaluation of its potential in a real-life context is still a questionable issue.

In this study, we investigate the behavior of a SMT engine faced with a corpus far different from the one it has been trained on. We show that terminological databases are obvious resources that should be used to boost the performance of a statistical engine. We propose and evaluate a way of integrating terminology into a SMT engine which yields a significant reduction in word error rate.

# 1 Introduction

La traduction statistique est devenue populaire au sein de la communauté langagière suite aux travaux de (Brown *et al.*, 1993). Depuis, de nombreux chercheurs se sont attelés à la réalisation de meilleurs modèles, et plusieurs approches séduisantes ont été proposées.

Même si les études sur la traduction statistique incluent habituellement une section d'évaluation fournie, il reste cependant difficile de savoir ce qu'on est en droit d'attendre des performances d'un moteur de traduction statistique sur une tâche donnée<sup>1</sup>. Nous savons des travaux de (Wang, 1998) que dans une tâche de traduction de parole, le moteur de traduction qu'il a développé se comparait favorablement à un système symbolique développé par plusieurs de ses collègues. Il est néanmoins hasardeux de généraliser ce constat à d'autres applications.

Nous ne connaissons pas d'étude systématique tentant d'évaluer l'adéquation de la solution statistique dans des *environnements de traduction réels*; terme que nous préférons d'ailleurs laisser sans définition. Il nous semble cependant évident qu'un système de traduction (statistique ou non) est d'autant plus viable à servir des applications variées qu'il est capable de s'adapter facilement à des textes d'une nature très différente de celle des corpus utilisés lors de la mise au point du système. Nous excluons donc de notre champ d'étude des systèmes hautement spécialisés qui par nature ne servent qu'une application très ciblée et peu évolutive, comme par exemple les systèmes METEO et ALT/Flash<sup>2</sup>. Curieusement, nous ne connaissons aucune étude sur l'*adaptabilité* des moteurs de traduction statistiques, et ce en dépit de l'abondante littérature décrivant des modèles de langue statistiques et adaptatifs.

Dans ce travail, nous évaluons les performances d'un moteur de traduction statistique lorsqu'il traduit des textes relevant de domaines très pointus, c'est à dire, très différents des corpus utilisés pour l'entraînement des modèles de langue et de traduction sous-jacents. Nous décrivons tout d'abord en section 2 notre engin de traduction. Nous quantifions ensuite en section 3 la baisse de performance d'un engin entraîné sur un corpus "général" (le HANSARD) lorsqu'on s'en sert pour traduire un texte très spécifique (dans cette étude, un manuel militaire pour les tireurs d'élite). Nous proposons alors en section 4 une méthode simple et naturelle d'améliorer un système de traduction général; à savoir, son ouverture à des ressources terminologiques disponibles. Nous montrons en section 5 les performances obtenues en implémentant notre approche et discutons en section 6 d'autres travaux auxquels la présente étude est liée.

## 2 Le moteur statistique

### 2.1 Les modèles statistiques

Pour ce travail, nous avons réalisé un engin traduisant du français vers l'anglais qui adopte le paradigme du canal bruité, initialement présenté dans le cadre de la traduction dans (Brown *et al.*, 1993) et qui peut se décrire simplement par l'équation 1, où  $e_1^J$  représente la séquence de mots cibles (ici des mots anglais) à trouver, étant donnée la phrase à traduire (ici des mots français) de  $J$  mots  $f_1^J$ .

<sup>1</sup>Le même constat peut d'ailleurs être fait à l'égard des systèmes non probabilistes.

<sup>2</sup>Je tiens à remercier mes relecteurs pour leurs commentaires avisés.

$$e_1^I = \operatorname{argmax}_{I, e_1^I} \underbrace{P(e_1^I)}_{\text{langage}} \cdot \underbrace{P(f_1^J | e_1^I)}_{\text{traduction}} \quad (1)$$

Pour entraîner les modèles probabilistes sous-jacents, nous avons réuni un *bitexte* de 1 639 250 paires de phrases du HANSARD alignées automatiquement au niveau des phrases. Dans cette expérience, tous les mots ont été convertis en lettres minuscules.

Nous avons utilisé un modèle trigramme interpolé entraîné sur la partie anglaise de notre bitexte. La perplexité du modèle résultant est assez basse – 65 – ce qui reflète les nombreuses formes figées présentes dans le HANSARD (ex: *pursuant to standing order / conformément à l'alinéa*).

Le modèle de traduction inversé (de l'anglais vers le français) utilisé ici est similaire au modèle 2 décrit dans (Brown *et al.*, 1993). Dix itérations du processus d'estimation des paramètres du modèle 1 ont été lancées (réduisant la perplexité de 7 776 à 90), suivies de 10 itérations du processus d'estimation des paramètres du modèle 2 (pour une perplexité finale de 54). Nous avons également réduit le nombre de paramètres du modèle de transfert (voir équation 2), en appliquant un algorithme décrit par (Foster, 2000) qui sélectionne les paires de mots les plus intéressantes d'un modèle<sup>3</sup>.

Le modèle 2 met également en jeu un modèle de longueur tel que spécifié dans l'équation 2. Dans cette étude nous avons fait l'hypothèse que la longueur (comptée en mots) d'une phrase française, traduction d'une phrase anglaise était normalement distribuée.

$$p(f_1^J | e_1^I) = p(J|I) \prod_{j=1}^J \sum_{i=0}^I \underbrace{p(i|j, J, I)}_{\text{alignement}} \cdot \underbrace{p(f_j | e_i)}_{\text{transfert}} \quad (2)$$

## 2.2 L'algorithme de recherche de la meilleure traduction

Nous avons étendu à un modèle trigramme le décodeur proposé par (Nießen *et al.*, 1998). L'idée de cet algorithme est d'étendre progressivement (c'est à dire mot à mot) les hypothèses de traductions, tout en couvrant progressivement les positions de la chaîne source. Nous invitons le lecteur à lire la description exacte de la récursion sur laquelle est construite la recherche et proposons à la place en figure 1 une vue programmatique du décodeur. Une hypothèse dans cette recherche est complètement déterminée par quatre paramètres: les positions source et cible, la couverture source d'une hypothèse et la nature du mot cible à la position cible. De ce fait, l'espace peut-être codé par une matrice creuse de dimension 4; chaque item dans cet espace de recherche contenant des informations de chaînage arrière (*backtracking*) ainsi que le score de l'hypothèse associée.

Nous savons que de meilleurs modèles de traduction ont été proposés et systématiquement comparés entre eux (Och & Ney, 2000). Les performances que nous avons relevées avec notre décodeur sur notre corpus HANSARD (voir la section 3) sont cependant comparables à (voire meilleures que) celles publiées ailleurs sur le même type de corpus. Notre but étant avant tout de comparer les performances d'un traducteur statistique utilisé dans des conditions *amicales*

<sup>3</sup>Nous avons ainsi conservé 1 million de paramètres sur un total initial de 34 969 331 paramètres

ou au contraire *adverses*, il nous semble donc que le moteur que nous avons utilisé sert tout à fait notre cause comparative.

**Entrée:**  $f_1 \dots f_j \dots f_J$

Initialize the search space table  $Space$

Select a maximum target length:  $I_{max}$

Compute the active vocabulary

```

for all target position  $i = 1, 2, \dots, I_{max}$  do
  prune( $i - 1$ );
  for all alive hyp.  $h = Space(i, j, c, e)$  do
     $uv \leftarrow History(h)$ ;
     $zones \leftarrow FreeSrcPositions(h)$ ;
     $bestWords \leftarrow NBestTgtWords(uv)$ ;
    for all  $w$  in  $bestWords$  do
       $prob \leftarrow Score(h) + \log p(w|uv)$ ;
      setIfBetter( $i, j, c, b, prob, 0, j, v$ );
      for all free source position  $d$  do
         $s \leftarrow prob$ ;
        for all  $f \in [1, f_{max}] / d + f - 1$  is free do
           $s+ = \log a(i|d, J) + \log t(f_d|e_i)$ ;
          setIfBetter( $i, d, c + f, w, s, f, j, w$ );

```

$max_s \leftarrow -\infty$

```

for all  $i \in [1, I_{max}]$  do
  for all alive hyp.  $h = Space(i, j, c, e)$  do
     $s \leftarrow Score(h) + p(J|i)$ ;
    if ( $(c == J)$  and ( $s > max_s$ )) then
       $max_s \leftarrow s$ 
       $\langle max_i, max_j, max_e \rangle \leftarrow \langle i, j, e \rangle$ 
if ( $max_s \neq -\infty$ ) then
  Return  $Space(max_i, max_j, J, max_e)$ ;
else
  Failure

```

**Sortie:**  $e_1 \dots e_i \dots e_{max_i}$

Figure 1: Principe de base de notre décodeur

## 3 Performances du moteur de traduction

### 3.1 Corpus de test

Dans cette section nous mesurons l'impact du type de corpus sur la performance de notre système. Nous utilisons à cet effet les deux corpus que nous décrivons ci-après. Le premier corpus (nommément, HANSARD) est une collection de phrases extraites d'une partie du corpus

HANSARD non utilisée lors de l’entraînement. Nous n’avons utilisé aucune stratégie particulière pour sélectionner ces phrases de manière par exemple à ce qu’elles soient proches des textes d’entraînement.

Le second corpus (dans la suite SNIPER) est un extrait d’un manuel militaire sur l’entraînement et le déploiement des tireurs d’élite; manuel qui a fait l’objet d’une autre étude (Macklovitch, 1995). Ce corpus relève hautement du domaine militaire et poserait sans aucun doute de nombreux problèmes à tout moteur (statistique ou non) non adapté à ce type de texte. Les caractéristiques principales de nos deux corpus sont regroupées dans la table 1.

corpus	<i>nb</i>	$ long. $	SER	WER
HANSARD	1038	$\langle 16.2, 7.8 \rangle$	95.6	59.6
SNIPER	203	$\langle 20.8, 6.8 \rangle$	100	74.6

Table 1: Caractéristiques principales de nos corpus de test et performance de notre moteur de traduction (voir la section suivante pour une description des taux SER et WER).  $|long.|$  indique la longueur moyenne (comptée en mots) des phrases sources ainsi que l’écart type de cette distribution; *nb* est le nombre de phrases dans le corpus.

## 3.2 Performances du système

Dans cette étude, nous évaluons la performance de notre moteur de traduction en terme de taux d’erreur mesurés au niveau de la phrase (SER) et des mots (WER). Ces deux taux sont mesurés en référence à un oracle disponible du fait que les deux corpus ont été publiés dans les deux langues. Le premier taux mesure le pourcentage de phrases pour lesquelles la traduction n’était pas exactement celle de l’oracle, alors que le second taux est calculé par une distance de Levenstein qui comptabilise le nombre minimal d’opérations qu’il faut effectuer pour passer de la traduction produite à la traduction oracle. Les trois opérations considérées ici sont l’*insertion*, la *suppression* et la *substitution* qui reçoivent toutes le même poids.

Nous sommes conscients que ces mesures à elles seules ne sont pas garantes d’une évaluation de qualité, mais nous étions hésitant à recourir dans cette étude à des évaluations humaines, en suivant par exemple le protocole décrit dans (Wang, 1998). En fait, un regard rapide sur la dégradation des performances mesurée sur le corpus SNIPER est tellement criante (voir la table 1), qu’il nous a semblé inutile de passer par des évaluations humaines pour la mettre en relief.

D’après la table 1, nous observons que les taux d’erreur au niveau des mots sur HANSARD sont de l’ordre de 60% alors qu’il est de 74% sur le corpus SNIPER. Il est intéressant de noter qu’aucune traduction sur ce dernier corpus n’a été identique à la traduction de l’oracle.

Bien qu’indiquant clairement une dégradation, il est difficile d’apprécier ce que ces taux d’erreur signifient véritablement. La table 2 nous aide à mieux comprendre les valeurs prises par le WER<sup>4</sup>. Il convient de noter que les taux observés sur le corpus *hansard* sont légèrement inférieurs à ceux rapportés récemment par (Och *et al.*, 2001) sur un corpus de même type. Lors d’une étude comparative de différents systèmes de décodage, les auteurs ont observé un taux de WER de l’ordre de 69% sur un corpus de 250 phrases (d’au plus 14 mots) extraites du corpus HANSARD.

<sup>4</sup>Les séances de traduction au complet sont disponibles à l’adresse:  
[www.iro.umontreal.ca/~felipe/ResearchOutput/TALN2002](http://www.iro.umontreal.ca/~felipe/ResearchOutput/TALN2002)

SRC	cependant , il y a ici deux problèmes qui apparaissent .	
REF	however , there are two problems here .	
CAN	however , there are two problems emerging here .	11%
SRC	les limites des circonscriptions électorales	
REF	electoral boundaries	
CAN	the electoral boundaries	33%
SRC	nous sommes fiers de ces habitants de london et d' autres canadiens qui consacrent leur temps et leur énergie à bâtir un monde meilleur .	
REF	we are proud of these londoners and of other canadians who devote their time and energies to improving our world .	
CAN	we are proud of these people of london and other people spend their time and energy to build a better world	50%
SRC	le mois de la nutrition	
REF	nutrition month	
CAN	in the month of nutrition	80 %
SRC	quelle plus belle image peut on donner du canada ?	
REF	this is canada at its best .	
CAN	what more can be nice to canada ?	100%

Table 2: Exemples de traductions extraites du corpus HANSARD à différents niveaux de WER.

### 3.3 Analyse de la baisse de performance

Deux raisons majeures expliquent les piètres performances observées sur le corpus SNIPER: la présence de mots hors vocabulaire et la traduction erronée des nombreuses unités terminologiques présentes dans le corpus.

Sur le corpus SNIPER, 3.5% des mots sources (formes) et 6.5% des mots cibles sont en effet inconnus des modèles statistiques. 44% des phrases sources et 77% des phrases cibles contiennent au moins un mot inconnu. Sur le corpus HANSARD, le taux de mots inconnus est beaucoup plus faible: environ 0.5% des mots sources et cibles sont inconnus et seulement 5% des phrases sources contiennent au moins un mot inconnu.

De manière prévisible, la présence massive de mots inconnus a un impact direct sur les performances et en particulier, sur la couverture du vocabulaire actif à partir duquel les traductions sont construites. Sur le corpus SNIPER, on mesure une couverture du vocabulaire de la traduction oracle de l'ordre de 72% (0.5% des phrases cibles oracles sont totalement couvertes), tandis que cette couverture s'élève à 86% sur le corpus HANSARD (24% des traductions de l'oracle sont complètement couvertes).

Il est en revanche beaucoup plus difficile de quantifier l'impact de la présence d'une terminologie spécifique sur la qualité de la traduction. Cela demanderait pour le moins d'identifier tous les termes et leur traduction. Une évaluation indirecte de cet impact est cependant fournie dans la section 5 où nous montrons que l'introduction d'entrées terminologiques améliore de manière significative les performances du système. La table 3 montre quelques exemples de mauvaise traduction impliquant des termes spécifiques au corpus SNIPER.

terme source	traduction oracle	traduction automatique
âme	bore	heart
huile polyvalente	general purpose oil	oil polyvalente
chambre	chamber	house of common
tireur d'élite	sniper	issuer of elite
la longueur de la crosse	butt length	the length of the crosse

Table 3: Exemples de traduction erronée pour quelques termes du corpus SNIPER.

## 4 Intégration de ressources terminologiques non probabilistes

Plusieurs stratégies sont envisageables pour tenter d'améliorer la situation. En tout premier lieu, il est possible que nous ayons à disposition des corpus spécifiques d'un domaine particulier en taille suffisante pour que l'on puisse entraîner un modèle spécifique que l'on peut ensuite combiner avec le modèle "général".

Nous pourrions de manière plus réaliste tenter de développer un modèle de traduction adaptatif. Un modèle cache pourrait par exemple être utilisé en combinaison avec notre modèle trigramme statique en ce qui concerne la composante langagière de notre moteur. La réalisation d'une composante traductionnelle adaptative est cependant une entreprise beaucoup plus spéculative qui nécessiterait entre autre une localisation assez précise des erreurs produites dans des traductions antérieures. Nous savons, suite aux travaux réalisés au sein du groupe de travail ARCADE (Véronis & Langlais, 2000), que l'alignement fin de mots est une tâche difficile.

Une troisième option s'offre à nous: tirer profit de ressources terminologiques existantes, comme par exemple *Termium*<sup>5</sup>. En fait, une des premières tâches du traducteur est souvent celle de la recherche terminologique; étape souvent prise en charge dans les organismes de traduction par des traducteurs terminologues (Langlais *et al.*, 2001). Il semble donc naturel d'un point de vue utilisateur d'ouvrir un système de traduction à des ressources terminologiques existantes (ou lexiques terminologiques dans la suite).

Parce qu'il est peu vraisemblable que ces ressources terminologiques soient livrées avec des probabilités de traduction, nous préférons voir un lexique terminologique comme un faisceau de contraintes qui permet de réduire l'espace de recherche de notre moteur. Savoir par exemple que le terme *tireur d'élite* se traduit souvent par le terme *sniper*, nous permet d'imposer à notre décodeur ayant à traduire une phrase contenant le terme français, de trouver une traduction contenant le terme *sniper*. Seule la position de ce dernier terme est à déterminer par notre décodeur.

Dans notre implémentation, nous avons légèrement modifié l'algorithme décrit dans la figure 1 afin, 1) d'interdire à tout mot anglais non validé par le lexique terminologique d'être associé à un mot source appartenant à une unité terminologique source, et 2) d'ajouter à toute position cible une hypothèse liant le terme source à l'une de ses traductions telles qu'identifiées dans le lexique. La survie de ces hypothèses dépend des contraintes globales imposées par l'opération de maximisation (de l'équation 1) sur laquelle repose la recherche.

Le score associé à une unité terminologique cible  $e_i^{j'}$  lorsque liée à sa contrepartie source  $f_j^{i'}$  est donnée par l'équation 3 où  $k$  désigne un indice cible et  $a(\cdot)$  le modèle d'alignement intervenant

<sup>5</sup>Voir <http://www.termium.com> pour plus d'information.

dans l'équation 2. L'intuition de cette équation est que les modèles consultés, à savoir le modèle trigramme et le modèle d'alignement, possèdent tous les deux une information qui peut aider à décider de l'adéquation d'une extension en position cible  $i$ . Le premier modèle fournit la probabilité qu'un mot donné (inconnu ou pas<sup>6</sup>) suive les deux derniers d'une hypothèse que l'on étend, tandis que le second modèle a une idée (faible) de la position source qui devrait être associée à un mot cible donné. Nous espérons qu'en l'absence de meilleurs alliés (un modèle cache améliorerait certainement les choses) ce mécanisme suffise à lui seul à contrôler la place finale de l'unité terminologique cible dans la traduction.

$$\sum_{k \in [i, i']} \log p(e_k | e_{k-2} e_{k-1}) + \max_{l \in [j, j']} \log(a(k|l, J, I)) \quad (3)$$

## 5 Résultats

Nous avons utilisé trois lexiques terminologiques dont les caractéristiques sont résumées dans la table 5. Ils diffèrent essentiellement par le nombre d'entrées qu'ils contiennent. Le premier lexique (`sniper-1`) contient les 33 entrées qui ont été employées dans une étude sur la vérification automatique de la consistance terminologique dans des traductions (Macklovitch, 1995). Le deuxième (`sniper-2`) et troisième (`sniper-3`) lexiques contiennent ces mêmes entrées plus d'autres ajoutées manuellement après inspection incrémentale de notre corpus de test SNIPER<sup>7</sup>.

Comme le montre la table 5, l'introduction d'un lexique terminologique dans le processus de traduction diminue les taux d'erreur mesurés au niveau des phrases et des mots, et ce, même avec des lexiques peu couvrants. Avec le lexique `sniper-1` nous observons une réduction absolue de 9.6% et une réduction de 13.8% avec le lexique `sniper-3`. La table 4 propose deux exemples de traductions produites avec et sans l'aide de lexiques spécialisés.

Il est important de noter, que si les performances sont meilleures, tous les problèmes ne sont pas pour autant réglés. Une inspection systématique des traductions proposées par notre engin de traduction en conjonction avec `sniper-3`, montre que si la traduction est de meilleure qualité que lorsque l'engin est utilisé sans lexique, il n'en reste pas moins qu'elle est moins fidèle au texte source que ne le sont les traductions obtenues sur le corpus HANSARD: les mots inconnus sont toujours un obstacle.

## 6 Discussion

Dans cette étude, nous avons montré que la traduction de textes hautement spécialisés à l'aide d'un engin de traduction probabiliste général est une tâche périlleuse. Ceci suggère une stratégie adaptative. Parmi les scénarios adaptatifs possibles, nous avons montré que l'ouverture du moteur de traduction à des ressources terminologiques est une approche naturelle et payante qui permet d'assouplir le moteur de traduction.

Ce travail est relié en partie à une étude récente de (Marcu, 2001), où l'auteur s'est intéressé à

<sup>6</sup>Notre modèle trigramme a été entraîné pour nous fournir des paramètres du type  $p(\text{UNK} | w_i w_{i+1})$ .

<sup>7</sup>Ce qui correspond à ce que fait le terminologue lorsqu'il identifie des termes dans le texte à traduire.

Source	le <b>tireur d'élite</b> voit simultanément les <b>files croisés</b> et l' image ( l' objectif ) .
Target	the <b>sniper</b> sees the <b>crosshairs</b> and the image - target - at the same time .
avec	<i>the gunman being same son sit and picture of the hon. members : agreed .</i>
sans	<i>the <b>sniper</b> simultaneously see the <b>crosshairs</b> and the image (objective . )</i>
Source	contrôle de la <b>détente</b> .
Target	exercising <b>trigger</b> control .
avec	<i>the control of détente .</i>
sans	<i>control of the <b>trigger</b> .</i>

Table 4: Deux exemples de traduction *avec* et *sans* lexique terminologique. Les termes concernés par le lexique sont typographiés en gras.

lexique	<i>nb</i>	<i>couverture</i>	SER	WER
sniper-1	33	20/247	99	67.4
sniper-2	59	47/299	98	66.2
sniper-3	146	132/456	98	64.3

Table 5: Performances du moteur de traduction avec différents lexiques terminologiques. *nb* est le nombre d'entrées dans le lexique et *couverture* indique le nombre d'unités sources différentes de ce lexique qui sont également dans le texte source à traduire, ainsi que le nombre total de leurs occurrences.

l'unification des approches de traduction statistique et basée sur l'exemple. Plus précisément, l'auteur a dérivé automatiquement du corpus HANSARD ce qu'il appelle une mémoire de traduction; en fait, une liste de paires de séquences de mots sources et cibles qui sont en relation de traduction. Ces paires ont été extraites à l'aide d'un alignement de viterbi utilisant un modèle de traduction de type IBM-4 (Brown *et al.*, 1993) également entraîné sur le corpus HANSARD. Cette liste (probabilisée) de séquences était alors insérée à un décodeur afin d'améliorer les performances globales de traduction.

Ce que cette étude suggère, c'est qu'une liste d'équivalents bilingues automatiquement extraite d'un corpus utilisé également pour l'entraînement d'un modèle de traduction statistique peut améliorer la performance de l'engin de traduction sous-jacent; résultat très intéressant en soi. Nous avons mené une étude semblable dans le contexte différent du projet TRANSTYPE, avec des résultats bien moins concluants (Langlais *et al.*, 2000). Au delà des différences liées aux modèles de traduction employés dans ces deux études (nous utilisons seulement un modèle 2), ainsi qu'aux différentes métriques utilisées, nous pensons que la différence de performance observée dans ces deux études s'explique par la nature même des corpus de tests. Le corpus utilisé dans (Marcu, 2001) consistait en un ensemble de 500 phrases d'au plus 10 mots, alors que le corpus utilisé dans (Langlais *et al.*, 2000) était plus larges et plus diversifié.

L'étude que nous avons décrite ici s'apparente aux deux études susmentionnées, à l'exception du fait que nous ne nous occupons pas ici de l'extraction automatique d'équivalents bilingues. Ce choix est motivé par les deux raisons suivantes: s'il est possible d'extraire automatiquement des unités bilingues<sup>8</sup>, il est cependant difficile de statuer sur la nature terminologique de ces unités. Nous pensons de plus qu'il est souhaitable que le traducteur soit responsable de la qualité des

<sup>8</sup>Des listes d'équivalents automatiquement acquis lors des expériences décrites dans (Langlais *et al.*, 2000) sont consultables à l'adresse [www.iro.umontreal.ca/~felipe/ResearchOutput/ANLP2000](http://www.iro.umontreal.ca/~felipe/ResearchOutput/ANLP2000).

lexiques introduits dans le moteur de traduction, car ils constituent un des rares moyens dont il dispose pour garder un peu de contrôle sur la sortie automatique produite; un point que les traducteurs professionnels semblent apprécier (Langlais *et al.*, 2001).

En guise de remarque finale, nous souhaitons souligner que nous concevons cette étude comme un premier pas vers l'unification entre l'approche basée sur l'exemple et l'approche statistique<sup>9</sup>. Nous souscrivons donc pleinement à l'idée développée dans (Marcu, 2001). Bien sûr, la traduction à partir d'exemples peut fournir bien plus que la simple liste d'équivalents utilisée dans cette étude (nous pensons notamment aux patrons traductionnels). La stratégie que nous entrevoyons pour cette unification est cependant identique dans l'idée à celle décrite ici; à savoir intégrer des *contraintes* dans le problème de recherche de la meilleure traduction. L'extension de cette notion de contrainte à des chaînes de mots qui ne sont pas nécessairement des séquences adjacentes de mots, ni même des chaînes complètement instanciées (patrons à trous) fait partie des pistes que nous souhaitons étudier dans le futur.

## Références

- BROWN P. F., PIETRA S. A. D., PIETRA V. J. D. & MERCER R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, **19**(2), 263–311.
- FOSTER G. (2000). A Maximum Entropy / Minimum Divergence translation model. In *Proceedings of the 38th Annual Meeting of the ACL*, p. 37–44, Hong Kong.
- LANGLAIS P., FOSTER G. & LAPALME G. (2000). Unit completion for a computer-aided translation typing system. In *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP)*, p. 135–141, Seattle, Washington.
- LANGLAIS P., FOSTER G. & LAPALME G. (2001). Integrating bilingual lexicons in a probabilistic translation assistant. In *Proceedings of the 8th Machine Translation Summit*, p. 197–202, Santiago de Compostela, Galicia, Spain: IAMT.
- MACKLOVITCH E. (1995). *Can Terminological Consistency be Validated Automatically ?* Rapport interne, CITI/RALI, Montréal, Canada.
- MARCU D. (2001). Towards a unified approach to memory- and statistical-based machine translation. In *Proceedings of the 39th Annual Meeting of the ACL*, p. 378–385, Toulouse, France.
- NIESSEN S., VOGEL S., NEY H. & TILLMANN C. (1998). A dp based search algorithm for statistical machine translation. In *Proceedings of the 36th Annual Meeting of the ACL and the 17th COLING*, p. 960–966, Montréal, Canada.
- OCH F. J. & NEY H. (2000). A comparison of alignment models for statistical machine translation. In *Proceedings of the International Conference on Computational Linguistics (COLING) 2000*, p. 1086–1090, Saarbrücken, Luxembourg, Nancy.
- OCH F. J., UEFFING N. & NEY H. (2001). An efficient a\* search algorithm for statistical machine translation. In *Proceedings of the Workshop on Data Driven Machine Translation yielded at the 39th Annual Meeting of the ACL*, p. 55–62, Toulouse, France.
- VÉRONIS J. & LANGLAIS P. (2000). *Evaluation of parallel text alignment systems: The ARCADE project*, volume 13, chapter 19, p. 369–388. Parallel Text Processing, Kluwer.
- WANG Y.-Y. (1998). *Grammar Inference and Statistical Machine Translation*. PhD thesis, CMU-LTI, Carnegie Mellon University.

---

<sup>9</sup>En fait, les deux approches sont basées sur l'exemple mais nous reprenons ici la terminologie anglophone (EBMT versus SMT).