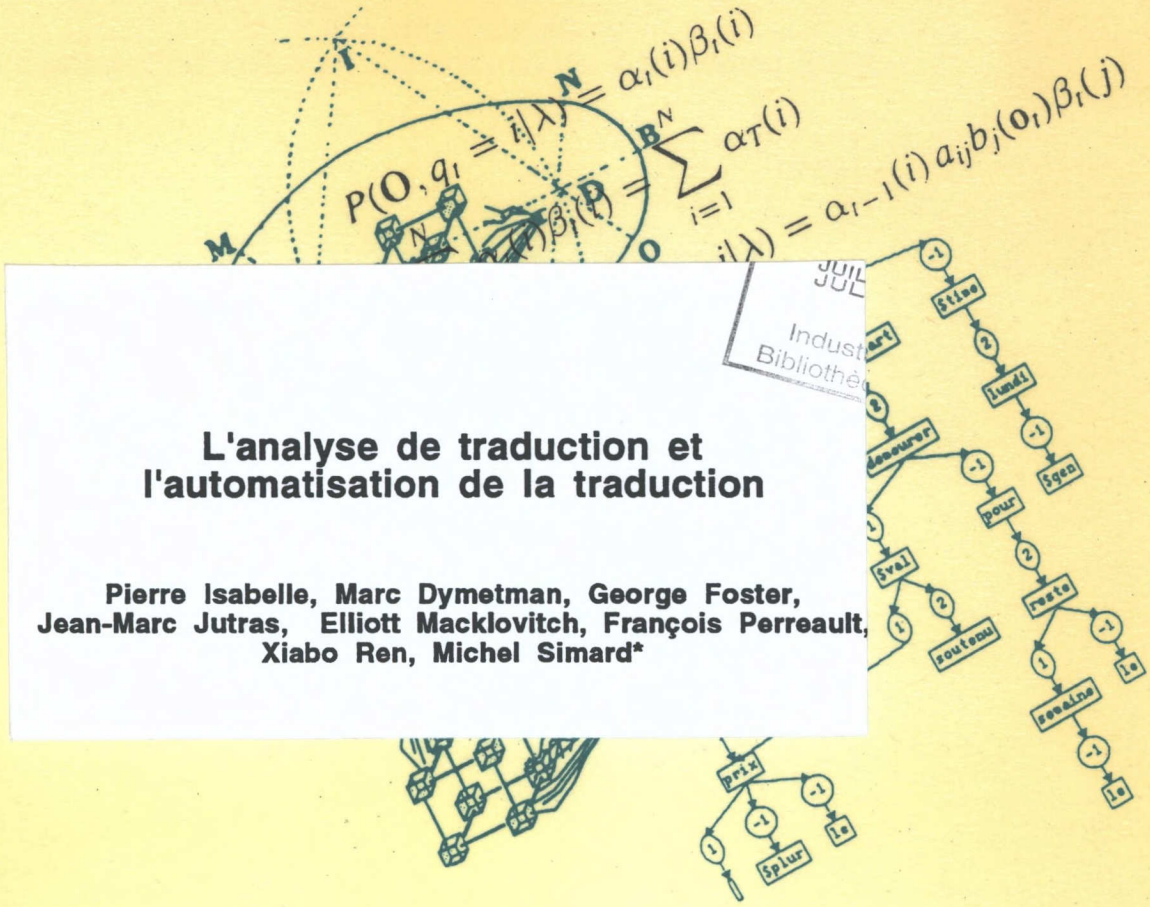
Industrie et Sciences Canada
Industry and Science Canada



L'analyse de traduction et l'automatisation de la traduction

Pierre Isabelle, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliott Macklovitch, François Perreault, Xiabo Ren, Michel Simard*

Industr
Bibliothè

CITI

Centre d'innovation
en technologies de l'information

Centre for Information
Technologies Innovation

QUEEN
P
98
.I814
1993
c.2

Canada

Queen
P98
.I8
1993
c.2
JOUR

Industrie Canada
Centre d'innovation en technologies de l'information (CITI)

Industry Canada
Library Queen
JUL 23 1998
Industrie Canada
Bibliothèque Queen

L'analyse de traduction et l'automatisation de la traduction

Pierre Isabelle, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliott Macklovitch, François Perreault, Xiabo Ren, Michel Simard*

Également publié en version anglaise dans les actes de la conférence «Fifth International Conference on Theoretical and Methodological Issues in Machine Translation», Kyoto, 1993.

Laval
Octobre 1993

* Courrier électronique : X@condor.citi.doc.ca, où X correspond à : isabelle, dymetman, foster, jutras, macklovi, perrault, ren ou simard.

Ce document fait état de travaux de recherche réalisés dans le cadre des activités du Centre d'innovations en technologies de l'information (CITI). Les opinions exprimées dans ce document n'engagent que l'auteur.

This report is also available in English.

© Copyright Industrie Canada 1994
No catalogue Co28-1/111-1994F
No ISBN 0-662-98969-4

RÉSUMÉ

Nous avançons que le concept d'*analyse de traductions* peut servir de point de départ à une nouvelle génération d'aides à la traduction. Nous montrons que les traductions peuvent être analysées et versées dans une *mémoire traductionnelle structurée* et décrivons le concordancier bilingue TransSearch que nous avons mis au point pour permettre aux traducteurs d'exploiter cette mémoire traductionnelle. Nous affirmons que les analyseurs de traductions peuvent contribuer à la détection des *erreurs de traduction* dans les premiers jets et nous présentons les résultats d'une expérience portant sur la détection des faux amis, réalisée dans le cadre du projet TransCheck. Nous soutenons enfin que l'analyse de traductions peut faciliter la transcription directe de traductions dictées et présentons le nouveau projet TransTalk.

1. Introduction

En 1951, Y. Bar-Hillel, qui fut le premier chercheur à se consacrer à temps plein au domaine de la traduction automatique (TA), écrivait :

> «Dans le cas des domaines cibles où la précision absolue est une condition *sine qua non*, il faut renoncer à la TA pure en faveur de la TA mixte, c'est-à-dire un processus traductionnel faisant intervenir l'intelligence humaine. Ce qui soulève la question suivante : Quelles étapes du processus devrait-on confier à un partenaire humain?»(Bar-Hillel [1], p. 230)

Quarante-deux ans et trois «générations» de systèmes plus tard, la TA pure n'est pas plus généralement applicable qu'elle ne l'était à cette époque[1]. Plus décourageant encore, la TA mixte ne l'est guère plus. Il n'existe pas de données précises aisément accessibles à ce sujet, mais on peut dire à coup sûr que la part actuelle de la TA, pure ou mixte, se situe bien en deça de 1 p. cent du marché global de la traduction. Force nous est donc d'en conclure que les chercheurs en TA n'ont pas encore réussi à proposer de réponses réalistes et pratiques à la question posée par Bar-Hillel au sujet de la division du travail entre l'homme et la machine.

---

1 Même s'il a été démontré que la TA peut être remarquablement performante dans le cas plutôt marginal de certains sous-langages extrêmement restreints, comme celui des bulletins météorologiques (Isabelle [12]).

Bar-Hillel lui-même proposa l'idée d'un tandem homme-machine dans lequel le partenaire humain interviendrait soit avant, soit après le processus mécanique, «mais de préférence pas quelque part au milieu». C'est donc dire que la machine se chargerait de l'essentiel du processus traductionnel. Depuis, la «TA assistée par l'hommme» est restée le paradigme prédominant dans le milieu de la TA, où l'on a continué à demander aux machines d'accomplir une tâche qu'elles n'arrivent pas àbien faire, c'est-à-dire traduire. Et où l'on a continué à demander aux traducteurs d'exécuter des tâches auxquelles ils préféreraient se soustraire, comme insérer des codes bizarres dans des textes sources, répondre à des questions inattendues sur la parenthésage des syntagmes ou réorganiser d'étranges fouillis de mots en langue cible. Résultat : le marché n'a jamais manifesté beaucoup d'enthousiasme pour ce genre de *modus vivendi* homme-machine.

Il est devenu évident qu'en général, les **machines ne réussissent toujours pas à maîtriser la partie essentielle du processus de traduction.** Déjà, en 1980, Martin Kay [18] plaidait en faveur d'un renversement des rôles qui aurait pour effet de remettre la machine à «sa place», c'est-à-dire celle d'assistant du traducteur humain :

> «Je veux préconiser une approche au problème selon laquelle on permettrait à la machine de prendre en charge graduellement, presque imperceptiblement, certaines fonctions du processus général de traduction. La machine assumerait d'abord des tâches non essentiellement reliées à la traduction. Puis, peu à peu, elle s'attaquerait à la traduction comme telle. Tout serait une question de modestie. À chaque étape, on ne confierait à la machine que ce qu'elle sait bien faire. Et ainsi, petit à petit, l'oiseau ferait son nid!»(p. 11)

C'est précisément pour cette approche réaliste et modeste que le Centre canadien de recherche sur l'informatisation du travail (qui porte maintenant le nom de Centre d'innovation en technologies de l'information - CITI) a opté, en 1987, lorsqu'il a lancé le projet de poste de travail de traducteur (Macklovitch [17]). Dans sa plus récente incarnation, le poste de travail du CITI met à la disposition du traducteur un environnement multi-fenêtres qui lui donne simultanément accès à plusieurs outils, comme le traitement de texte en écran partagé, la vérification orthographique, la consultation terminologique et lexicographique, la comparaison de fichiers, le compte de mots, l'extraction de textes intégraux, etc. (Macklovitch [17]). Il faut reconnaître que ces fonctions ont plus à voir avec la bureautique qu'avec l'automatisation de la traduction elle-même. Mais suivant le scénario proposé par Kay, nous pouvons maintenant tirer profit de cette base informatique en l'enrichissant progressivement d'outils orientés vers la traduction. Dans cette optique, la question centrale peut être formulée comme suit : **au-delà de la bureautique intégrée, mais en deça de la**

traduction automatique, que peut-on faire de plus pour faciliter la tâche du traducteur?

Nous avançons, dans la suite de cet article, que le concept d'*analyse de traductions* peut servir de fondement à l'élaboration d'une nouvelle génération d'outils informatisés à l'intention du traducteur. La deuxième section de cet article est consacrée à une description générale de la notion d'analyse de traductions. Les sections 3, 4 et 5 décrivent les travaux réalisés au CITI sur trois applications : une mémoire traductionnelle, un vérificateur de traductions et un système de dictée pour traducteurs.

2. L'analyse de traductions

Dans la documentation récente (Dymetman et Macklovitch [14], par exemple), la traduction est souvent conceptualisée comme une relation $tr_{L1L2}(s, t)$, dont l'extension est un ensemble de paires $<S, T>$, où S est un texte en langue L_1 et T un texte en langue L_2. Comme il existe, dans chacune de ces langues, un nombre infini de textes, la relation tr_{L1L2} doit être définie de façon récursive, ce qui aura pour conséquence que cette relation aura un caractère compositionnel : jusqu'au niveau d'un certain ensemble fini d'éléments élémentaires, S et T seront décomposés respectivement en des ensembles d'éléments $\{s_1, s_2, \ldots, s_n\}$ et $\{t_1, t_2, \ldots, t_n\}$, de telle façon que pour tout i, la relation $tr_{L1L2}(S_i, T_i)$ sera également satisfaite.

Un système de TA ordinaire incorpore une spécification quelconque (éventuellement partielle) de la relation traductionnelle tr_{L1L2}, de même qu'un procédure qui produira, pour n'importe quelle valeur de S, une ou plusieurs valeurs T pour lesquelles $<S, T>$ appartiendra à tr_{L1L2}. Un système de TA réversible (voir, par exemple, Dymetman [8], Van Noord [21]) peut en outre calculer, pour n'importe quelle valeur de T, les valeurs S pour lesquelles $<S, T>$ appartiendra à tr_{L1L2}.

Les systèmes de TA tentent de résoudre le problème de la production de traductions, mais, comme l'a fait remarquer Debili [7], nous pouvons aussi envisager les traductions du point de vue de la reconnaissance. Nous appellerons donc *accepteur de traductions* toute procédure qui, à partir d'une paire particulière de textes $<S, T>$, peut déterminer si la relation $tr_{L1L2}(S, T)$ est toujours vérifiée. Et nous appellerons *analyseur de traductions* toute procédure récursive $at(<S, T>, AAT)$ qui attribue, aux paires $<S, T>$ qui satisfont la relation $tr_{L1L2}(S, T)$, un arbre d'analyse traductionnelle

AAT. Un AAT rend explicite la structure compositionnelle de la relation traductionnelle. Par exemple, moyennant une définition appropriée de la relation traductionnelle anglais-français, un analyseur de traductions pourrait produire un AAT comme celui qui est illustré à la Figure 1.

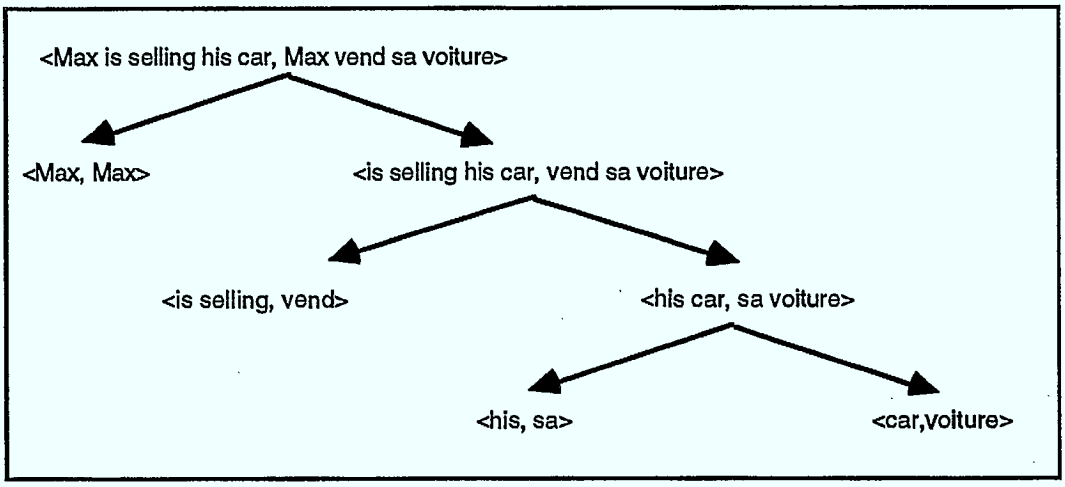


Figure 1 : Un arbre d'analyse traductionnelle (AAT)

Isabelle [13] utilise le terme *bi-texte* pour désigner une structure qui, comme l'AAT, sert à décomposer les traductions en leurs correspondances constitutives. Les AAT sont des descripteurs structuraux des analyses de traductions au même titre que les arbres d'analyse ou de parsage sont des descripteurs structuraux des analyses grammaticales.

L'analyse de traductions et la TA posent, en principe, des problèmes très semblables : le calcul est basé sur la même relation abstraite tr_{L1L2} La différence réside uniquement dans les modes de calcul. Cela signifie-t-il que les analyseurs de traductions et les systèmes de TA sont, en pratique, assujettis exactement aux mêmes contraintes? Plus spécifiquement, cela signifie-t-il qu'il n'est possible de réaliser des analyseurs de traductions efficaces qu'à la seule condition qu'il soit également possible de réaliser des systèmes de TA efficaces?

Il est évident que non. Dans les rares cas où la TA de haute qualité est possible, il devrait évidemment être possible de construire un analyseur de traductions pour les sorties du système de TA. Dans les cas où la TA n'est pas possible, nous soutenons, et c'est ce qui importe, qu'il est malgré tout possible d'élaborer des dispositifs capables d'analyser les traductions réalisées par des humains et que ces analyseurs auront de nombreuses utilités. Cette différence découle des

exigences pratiques que des tâches différentes (la TA par opposition à l'analyse de traductions) imposent au niveau de précision de la caractérisation formelle de la relation $tr_{L1L2}(S, T)$.

Considérons, par exemple, le modèle qui sous-tend la méthode d'alignement de phrases proposée par Brown *et al.* [3]. Sur le plan conceptuel, ce modèle génère des séquences de paires de textes $<S, T>$ qui présentent les caractéristiques suivantes : a) S est une séquence $<s_1, s_2, ..., s_n>$ dans laquelle chaque s_i est lui-même une séquence de zéro, une ou deux «phrases» et T est une séquence semblable $<t_1, t_2, ..., t_n>$; b) une «phrase» est une chaîne d'unités lexicales terminée par une unité de ponctuation; c) une unité lexicale est une chaîne de caractères encadrée par des caractères délimitateurs; d) la longueur $l(s_i)$ de chaque s_i (en termes du nombre d'unités contenues) présente une corrélation avec la longueur $l(t_i)$ de la «phrase» correspondante t_i, selon une distribution de probabilités $pr(l(s) \mid l(t))$ et e) cette distribution de probabilités peut être estimée à partir des fréquences observées dans des corpus de traductions, comme le corpus bilingue du Journal des débats de la Chambre des communes.

Ce modèle saisit bien l'un des aspects spécifiques de la relation traductionnelle établie entre deux langues, à savoir la corrélation de longueur qui existe entre les phrases qui sont des traductions réciproques. En ce sens, il constitue un modèle de traduction, aussi faible soit-il.

Si l'on appliquait un modèle de ce genre à la traduction anglais-français, une phrase anglaise e se traduirait plus ou moins par une séquence aléatoire de caractères f, dont la seule propriété notable serait d'avoir une longueur $l(f)$ qui est typique de la traduction française d'un phrase anglaise de longueur $l(e)$. Dans la pratique, un tel «système de TA» semblerait parfaitement inutile.

Si, d'autre part, à l'instar de Brown *et al.*, on applique ce modèle à l'analyse de traductions, on obtient un système capable de décomposer des traductions existantes en des représentations qui rendent leur structure compositionnelle explicite jusqu'au niveau de la phrase. Le résultat sera un arbre d'analyse ayant la forme illustrée à la Figure 2, où les textes S et T sont décomposés en n paires successives de blocs de phrases s_i et t_i.

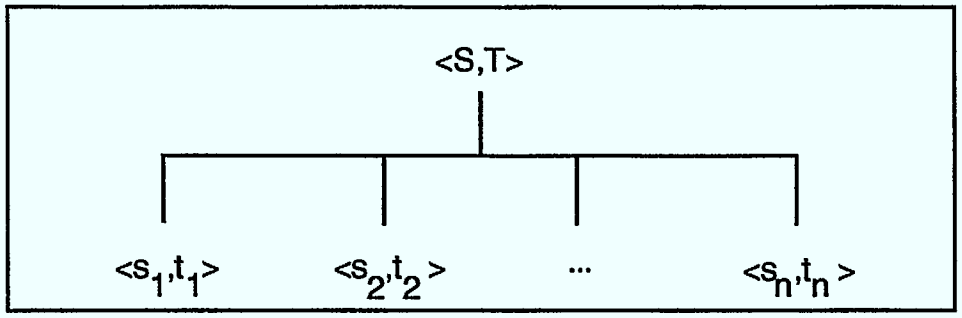


Figure 2 : L'alignement de phrases en tant qu'arbre d'analyse simple

L'analyse est, à l'évidence, très grossière : aucune correspondance n'est établie au-delà du niveau de la phrase. Pourtant, comme nous le verrons bientôt, ces bi-textes à «faible résolution» peuvent servir de base à certains outils informatisés très pratiques.

Évidemment, des analyses plus raffinées offriraient encore plus de possibilités à cet égard. En fait, il n'est pas trop difficile d'imaginer des familles de modèles de traduction un peu plus puissants qui, tout en demeurant insuffisants pour les fins de la TA, pourraient néanmoins être utilisés pour expliciter davantage de structures dans des traductions existantes (comme les correspondances entre syntagmes ou mots).

Pour ce qui est de leur architecture générale, les modèles utilisés pour l'analyse de traductions peuvent être très proches de ceux utilisés pour la TA. La possibilité la plus évidente est peut-être le modèle tripartite illustré à la Figure 3.

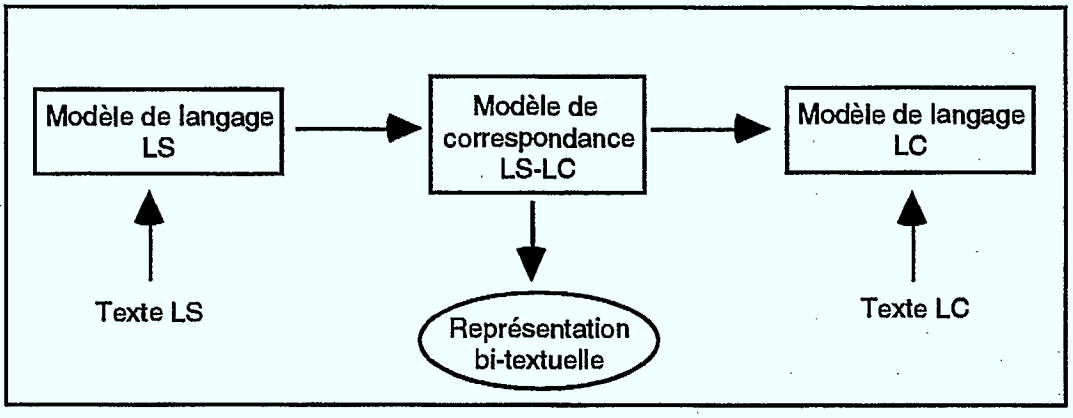


Figure 3 : Un modèle tripartite pour l'analyse de traductions

Comme le modèle de transfert bien connu de la TA, ce modèle comprend deux composantes propres à chacune des langues (les modèles de langage) et une composante «contrastive», propre aux paires (le modèle de correspondance). Les deux composantes unilingues fonctionnent en mode analytique et les représentations linguistiques qu'elles produisent sont traitées par le modèle de correspondance, qui les relie en une représentation bi-textuelle unique où les correspondances traductionnelles sont rendues explicites. Que ses composantes soient mises en oeuvre au moyen de techniques à base de règles ou à base de corpus, ce modèle demeure un modèle naturel. En fait, la meilleure façon de conceptualiser la méthode d'alignement simple basée sur la longueur des phrases, que nous avons déjà mentionnée, est de l'envisager comme une instance de ce modèle.

Certaines indications suggèrent que les modèles probabilistes se révéleront extrêmement utiles pour l'élaboration d'analyseurs de traductions d'utilité générale. Tandis que les méthodes à base de règles conviennent bien à l'élaboration de modèles «profonds» dans des domaines restreints, les méthodes probabilistes semblent particulièrement bien adaptées au développement de modèles superficiels potentiellement capables de produire des analyses partielles relativement précises de traductions générales.

Quoi qu'il en soit, l'essentiel de notre propos est que l'analyse de traductions, même basée sur des modèles de traduction faibles, constitue un point de départ approprié pour le développement d'une nouvelle génération d'aides à la traduction. Nous examinerons maintenant certains de ces outils.

3. La mémoire traductionnelle

3.1 Les traductions existantes en tant que ressource

La tendance aux approches à base de corpus en TA découle en partie de la constatation que le fonds de traductions existantes constitue une ressource d'une très grande richesse dont le potentiel n'a pas encore été pleinement exploité. En fait, il est évident que les **traductions existantes renferment infiniment plus de solutions à plus de problèmes de traduction que tout autre outil de référence.**

Mais les traducteurs ne pourront exploiter les richesses enfouies dans leurs traductions antérieures que lorsqu'ils disposeront des outils leur permettant de les gérer comme des données de traduction plutôt que comme des données de traitement de texte. Et c'est précisément à cela que sert un analyseur de traductions : transformer des données de traitement de texte en des structures bi-textuelles qui rendent explicites les correspondances traductionnelles.

Une fois les traductions existantes structurées en bi-textes, les segments correspondants en langue source et en langue cible sont systématiquement interreliés. Plus particulièrement, tout segment qui renferme une occurrence d'un problème de traduction est relié à un segment qui renferme une solution toute faite à ce problème. S'ils disposent des moyens nécessaires pour créer, stocker et interroger de telles structures bi-textuelles, les traducteurs pourront transformer leur production antérieure en une mémoire traductionnelle exploitable et extrêmement efficace.

3.2 TransBase

Pour rendre accessibles les résultats des analyses traductionnelles de grandes quantités de texte, nous avons conçu un modèle simple de mémoire traductionnelle structurée, que nous avons appelé TransBase. Ce modèle possède les mêmes caractéristiques de base que les systèmes d'extraction de textes intégraux : il peut gérer des quantités arbitraires de texte, il peut être augmenté de façon incrémentielle et il assure un accès rapide au contenu textuel de la base de données. Ce qui le distingue essentiellement de ces systèmes, c'est sa capacité à stocker des représentations bi-textuelles.

Une base de données TransBase se construit à l'aide d'un analyseur de traductions semblable à celui qui est illustré à la Figure 2. Chacun des textes d'une paire de traductions réciproques fait l'objet d'une analyse linguistique qui le décompose en ses éléments structuraux (paragraphes, phrases, etc.) et détermine son contenu lexical. Cette information est stockée dans deux composantes distinctes de la base de données, propres à chacune des langues en présence, et indexée de façon à permettre l'accès rapide à n'importe quelle partie du texte. Un «analyseur de correspondances», basé sur les techniques décrites dans Simard, Foster et Isabelle [20], utilise ensuite ces analyses linguistiques pour construire une «carte traductionnelle» au niveau de la phrase, qui est également stockée dans la base de données. Cette structure et le mode de construction de la base de données sont illustrés à la Figure 4.

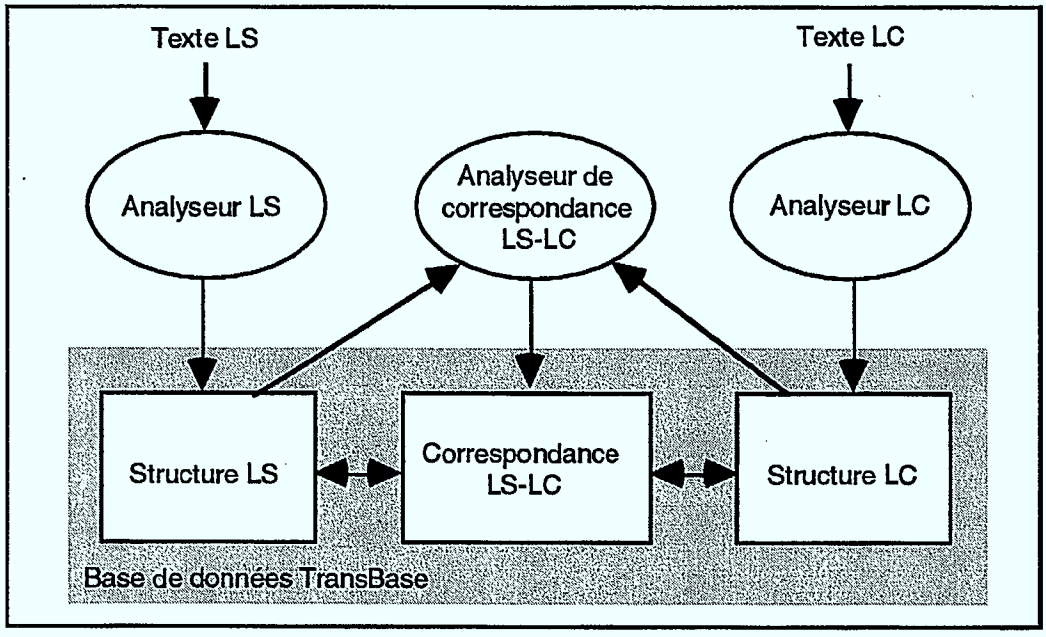


Figure 4 : Structure générale de la base de données TransBase

Dans la base de données, les textes source et cible sont traités de façon symétrique. Cependant, comme la directionalité de la traduction peut être importante pour l'utilisateur, TransBase peut enregistrer laquelle des deux langues est la langue source.

3.3 TransSearch

Il existe de nombreuses façons d'exploiter cette mémoire traductionnelle. La première qui vient à l'esprit, et celle qui est probablement la plus universellement utile, consiste à fournir aux traducteurs des outils qui leur permettront d'interroger le contenu textuel de la base de données TransBase. Certains auteurs (voir, par exemple, Church et Gale [6]) ont déjà suggéré qu'un outil capable de produire des *concordances bilingues* serait utile aux lexicographes bilingues. Il est plutôt évident qu'un concordancier bilingue serait également utile aux traducteurs. Il se peut, par exemple, qu'en rencontrant, dans un texte de départ anglais, un idiotisme tel que *to be out to lunch* ou *to add insult to injury*, le traducteur ne soit pas certain de l'équivalent français approprié. Il se peut aussi qu'il ne trouve pas de réponse satisfaisante dans les dictionnaires bilingues traditionnels. S'il disposait d'un concordancier bilingue, il pourrait alors interroger une base de données bi-textuelles du genre de TransBase et en extraire des exemples de ces expressions, accompagnées de leurs traductions. Cela serait pratique non seulement pour les idiotismes, mais aussi pour la terminologie spécialisée

ou pour les tournures et formules propres à certains domaines (*To whom it may concern..., Attendu que...).* On trouvera, dans Macklovitch [16], un exposé plus détaillé de cette question.

Le logiciel TransSearch est justement cet outil : il permet d'extraire de la base de données des occurrences d'«expressions» précises et de les afficher à l'intérieur de leur contexte bilingue. Étant destiné principalement aux traducteurs, qui sont susceptibles de l'utiliser simplement comme source de référence supplémentaire, ce logiciel est conçu pour être exploité en mode interactif et pour donner des réponses en temps réel, ce qui est assuré par l'inclusion d'index de mots dans le modèle TransBase.

Comme la plupart des traducteurs ne sont pas des experts en informatique, nous avons accordé beaucoup d'attention à la convivialité de l'interface de TransSearch. En utilisant un langage d'interrogation intuitif et à orientation graphique, il est facile pour l'utilisateur d'effectuer des recherches complexes dans la base de données. Chacune de ces interrogations définit une expression logique portant sur des séquences de mots : lorsque l'interrogation est lancée, le système extrait de la composante alignement de la base de données tous les couples qui correspondent à cette expression. L'inclusion de dictionnaires et de descriptions morphologiques du français et de l'anglais permet en outre à TransSearch de repérer automatiquement les formes fléchies des éléments recherchés.

Les résultats d'une interrogation sont normalement présentés en deux colonnes, les traductions réciproques étant affichées côte-à-côte. L'utilisateur peut alors examiner chacune des solutions proposées à l'intérieur du document dont elle a été extraite, ou rassembler toutes les solutions repérées accompagnées d'une petite portion de leur contexte immédiat, selon le mode de présentation habituel des concordances.

La Figure 5 présente un exemple type des résultats obtenus avec TransSearch. Dans cet exemple, l'utilisateur a interrogé le système pour trouver des occurrences de l'expression anglaise *to take X to court* qui ne sont pas traduites en français par *poursuivre X* ou *intenter un (ou des) procès à X* et la base de données interrogée était constituée des traductions du Journal des débats de la Chambre des communes de 1986. Tous les traducteurs à qui nous avons montré le fonctionnement de ce système en ont conclu qu'un concordancier bilingue leur serait très utile.

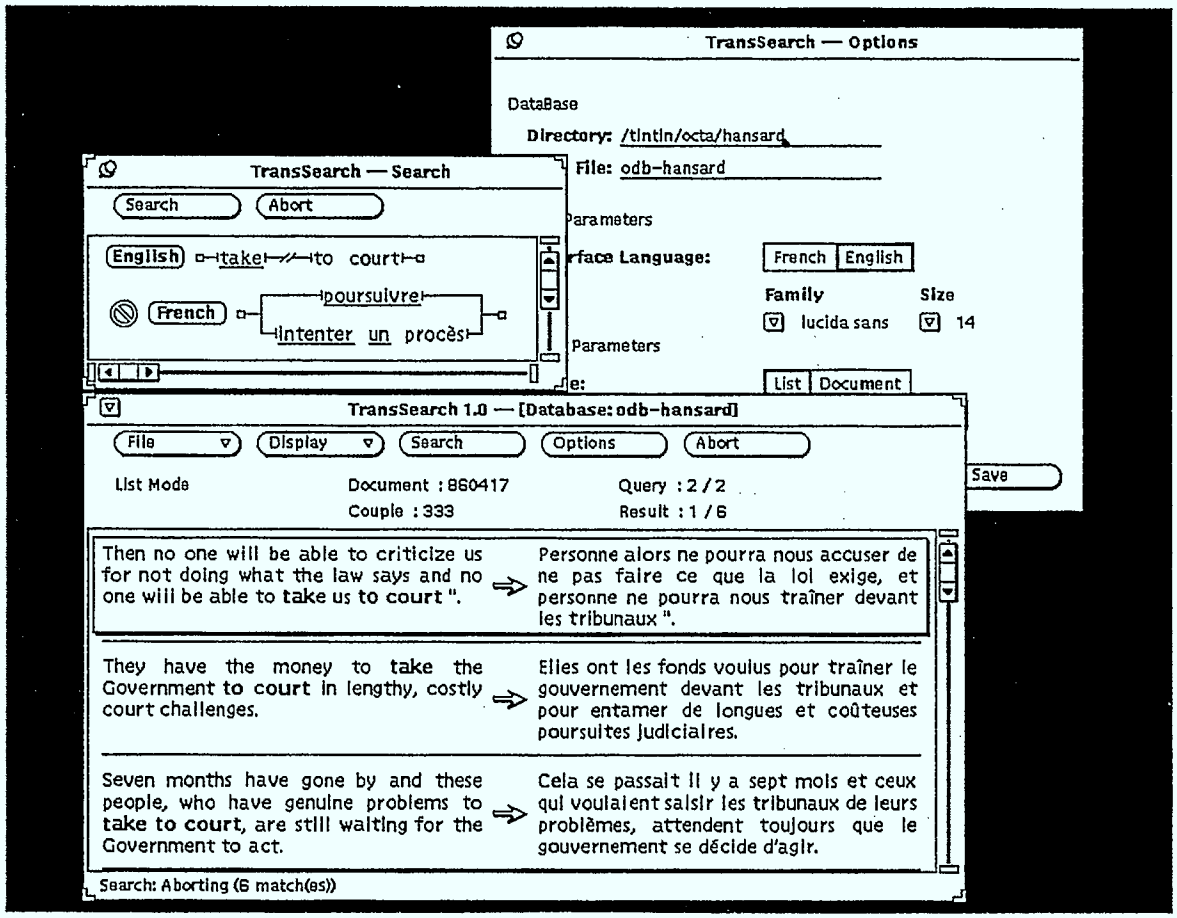


Figure 5 : Exemple des résultats obtenus avec TransSearch

4. La vérification de traductions

4.1 L'analyse de traductions et la détection d'erreurs

Depuis quelques années, on voit apparaître sur le marché du logiciel des outils critiques conçus pour aider les rédacteurs à améliorer leurs textes, grâce à la détection des problèmes potentiels d'orthographe, de grammaire et même de style. Si ces outils peuvent, en principe, aider les traducteurs à corriger les erreurs de rédaction contenues dans leurs traductions, ils ne peuvent aucunement les aider à corriger les *erreurs de traduction* au sens strict du terme, c'est-à-dire les correspondances erronées entre le texte source et le texte cible. Par exemple, ils ne peuvent pas

les aider à repérer les *contresens* ou les *faux sens*, c'est-à-dire les cas où les deux textes sont individuellement corrects et signifiants, mais ne sont pas, en l'occurrence, des traductions réciproques. De telles erreurs ne peuvent être détectées que par un dispositif qui examine simultanément le texte source et le texte cible : en d'autres termes, un dispositif qui incorpore un analyseur de traductions.

Moyennant un analyseur de traductions capable de reconstruire un sous-ensemble quelconque *Cset* des correspondances observables dans les résultats d'une opération de traduction, et moyennant un ensemble quelconque de contraintes *C* auxquelles doivent satisfaire les correspondances admissibles, un vérificateur de traductions est un dispositif qui aide le traducteur à s'assurer que le sous-ensemble *Cset* respecte effectivement les contraintes *C*. Cette capacité exige la présence d'un analyseur de traductions basé sur un modèle de traduction «robuste», un modèle qui est capable de détecter les correspondances réelles susceptibles de dévier de la norme définie par *C*.

Le problème général de la qualité des traductions est une question manifestement complexe et frustrante. Nous n'avons certainement pas l'intention de proposer une mesure ou une méthode globale pour évaluer les traductions. Notre but est plus modeste : nous désirons seulement cerner certaines caractéristiques particulièrement simples qui sont recherchées par la plupart des traducteurs et mettre au point quelques outils qui les aideront à vérifier si leurs traductions possèdent ces caractéristiques.

Une première caractéristique souhaitable et plutôt évidente est l'*exhaustivité*. En effet, toutes les parties d'un texte source devraient normalement avoir un équivalent dans le texte cible. Mais il arrive parfois aux traducteurs de faire des *erreurs d'omission*, en oubliant par exemple de traduire une phrase, un paragraphe, voire un page entière. Dans de tels cas, un analyseur de traductions efficace devrait être en mesure d'établir qu'un segment du texte source est mis en correspondance avec un segment vide dans le texte cible. Le dispositif de vérification pourrait alors en informer le traducteur, en lui signalant l'existence d'un problème éventuel dans son premier jet.

Une deuxième caractéristique candidate est la *cohérence ou l'uniformité terminologique*. En traduction technique, il est de rigueur d'utiliser systématiquement le même terme pour traduire toutes les occurrences d'un terme particulier du texte source. Un processus d'analyse permettant de faire ressortir toutes les correspondances terminologiques entre une traduction et sa source

devrait vraisemblablement aider les traducteurs à respecter le principe de l'uniformité terminologique.

Autre caractéristique : les traductions sont censées être exemptes *d'interférences linguistiques* provenant de la langue source. Dans certains cas, ces interférences mènent à des constructions boiteuses en langue cible, que l'on peut souvent détecter sans même se référer au texte source. Par exemple, si le mot anglais *address* est traduit en français par *addresse* (avec deux d), un vérificateur orthographique ordinaire devrait être mesure de détecter cette erreur. Mais il y a aussi des cas où l'interférence mène non pas à des mots mal orthographiés, mais bien à des erreurs de traduction. Les *faux amis*, par exemple, ont tendance à provoquer ce genre d'interférence.

Le mot m_e de la langue L_e et le mot m_f de la langue L_f sont des *mots apparentés* lorsque leur forme est semblable en raison d'une étymologie commune. C'est le cas, par exemple, du mot anglais *government* et du mot français *gouvernement*. Le plus souvent, ces mots sont non seulement des homonymes trans-linguistiques, mais aussi des synonymes. Dans certains cas, cependant, il n'y a pas de synonymie. Par exemple, les mots apparentés anglais/français suivants ont des sens complètement différents : *<actual/actuel>*, *<library/librairie>*, *<physician/physicien>*. Ces mots apparentés sont dits des «faux amis» parce que leur ressemblance morphologique crée une attente sémantique qui peut induire en erreur. La phrase *Max se rendit à la librairie* est parfaitement correcte en français, mais comme traduction de *Max went to the library*, elle constitue un cas flagrant d'erreur de traduction. Si un analyseur de traductions était capable de reconnaître, dans une traduction, une correspondance établie entre des mots apparentés reconnus comme des faux amis, il pourrait marquer cette correspondance comme une erreur possible que le traducteur pourrait ensuite vérifier.

Il existe probablement plusieurs autres types d'erreurs de traduction qu'un analyseur de traductions pourrait aider à déceler. La recherche dans ce domaine ne fait que commencer. Afin d'avoir un meilleur aperçu du potentiel pratique de cette approche, nous avons réalisé une expérience portant sur la détection des faux amis dans des traductions réelles.

4.2 Une expérience portant sur la détection des faux amis

Les faux amis (FA) peuvent être subdivisés en faux amis *absolus* ou *partiels*. Les FA absolus, comme ceux qui sont cités dans les exemples précédents, se caractérisent par le fait que leurs significations sont complètement disjointes; ils ne peuvent donc jamais être utilisés comme des traductions réciproques. Les FA partiels, par contre, ont des sens qui se recoupent partiellement et ils peuvent être des équivalents traductionnels dans un sous-ensemble particulier de leurs emplois possibles. Par exemple, le verbe français *examiner* est parfois l'équivalent (≡) et parfois le non-équivalent (≠) du verbe anglais *to examine* :

The doctor examined his patient ≡ *Le médecin examina son patient*
The professor examined his students ≠ *Le professeur examina ses étudiants*

En nous concentrant dans un premier temps sur le problème plus facile des FA absolus, nous avons réalisé une expérience visant à évaluer 1) l'ampleur du problème dans des traductions réelles et 2) l'efficacité de certaines méthodes de détection simples.

Nous avons élaboré un analyseur de traductions AT1 qui instancie le modèle de la Figure 1 de la façon suivante : les modèles de langage pour le français et l'anglais sont réduits à des processus de segmentation en mots (*tokenisation*) et d'analyse morphologique (basés sur un dictionnaire et sur un ensemble de règles de flexion). Ces modèles de langage produisent une représentation morphologique simple du texte d'entrée : chaque unité lexicale est représentée comme l'ensemble des formes des entrées lexicales dont elle est une instance possible. Le modèle de correspondance utilisé dans AT1 est simplement le programme d'alignement de phrases mis au point par Simard, Foster et Isabelle [20]. La représentation qu'il produit est une séquence $<<e_1, f_1>, <e_2, f_2>, ...<e_n, f_n>>$, dans laquelle chaque e_i est une séquence de zéro, une ou deux phrases du texte anglais représentées morphologiquement, chaque f_i est une séquence de zéro, une ou deux phrases du texte français représentées morphologiquement, et chaque $<e_i, f_i>$ est une correspondance traductionnelle.

Nous avons extrait de van Roey *et al.* [22] une liste de 145 paires de mots classifiés comme des faux amis absolus : *<accommodate/accommoder>*, *<actually, actuellement>*, etc.[2]. Nous avons ensuite réalisé un vérificateur simple qui aurait pour fonction d'examiner les résultats produits par AT1 et d'identifier, pour chaque paire de mots $<M_e, M_f>$, l'ensemble de paires de phrases $<e_i, f_i>$ satisfaisant la condition $M_e \in e_i$ (c'est-à-dire que e_i contient le mot m_e) et $M_f \in e_f$. Cette condition peut évidemment être satisfaite par des paires de phrases où m_e et m_f sont présents sans toutefois être utilisés comme des traductions réciproques.

Nous avons ensuite testé ce dispositif relativement simpliste sur un corpus composé des traductions du Journal des débats de la Chambre des communes couvrant une période d'un an. Après vérification manuelle des résultats, nous avons constaté qu'un grand nombre d'erreurs de traduction réelles avaient été détectées, comme dans l'exemple suivant :

> The peace movement in Canada is composed of **physicians**, members of the church, [...]
> _ Le mouvement canadien pour la paix compte dans ses rangs des **physiciens**, des ecclésiastiques, [...]
>
> (Journal des débats, 1987/09/29)

> There are parts of this bill which concern **librarians** and the artistic community.
> _ Quelque part dans ce projet de loi, il est question des **libraires** et des artistes.
>
> (Journal des débats, 1987/11/30)

Mais, comme on peut le voir dans le Tableau 1, les résultats présentaient aussi un niveau de «bruit» très élevé.

	Nombre de cas	Pourcentage
Erreurs réelles	57	7,4
Bruit	718	92,6
Total	775	100

Tableau 1 : Résultats de l'extraction de FA dans les sorties de AT1

2. Nous ne savons pas encore quelle proportion du problème réel est prise en compte par ces 145 paires, mais nous avons de fortes de raisons de soupçonner qu'il s'agit seulement de la pointe de l'iceberg.

Ce bruit provenait de trois sources différentes. Premièrement, il y avait des cas où la «fausseté» (par allusion à faux amis) de $<M_e, Mf>$ était attribuable à la catégorie grammaticale (POS) des deux mots en présence. Par exemple, le nom français *local* et le nom anglais *local* sont des faux amis absolus, mais leurs homographes adjectivaux n'en sont pas. Le vérificateur ne tenant pas compte de l'information grammaticale, il a relevé des cas non pertinents. Deuxièmement, une certaine proportion de bruit était générée par des citations non traduites. Par exemple, le mot anglais *agenda* et le mot français *agenda* sont des FA absolus. Comme ils sont parfaitement identiques, notre vérificateur a été incapable de les distinguer et a donc extrait des cas où le mot *agenda* apparaissait des deux côtés, pour la simple raison que l'un des textes le renfermait sous la forme d'une citation non traduite de l'autre langue. Troisièmement, il y avait des cas où M_e et M_f apparaissaient effectivement dans des phrases qui étaient des traductions réciproques, mais où ces mots n'étaient pas eux-mêmes utilisés comme des équivalents traductionnels. Notre modèle de correspondance (c'est-à-dire d'alignement de phrases) était simplement trop grossier pour éliminer ces cas[3]. Ces sources de bruit sont ventilées dans le Tableau 2.

Catégorie de bruit	Nombre de cas	Pourcentage
Mauvaise catégorie grammaticale	703	97,9
Citation non traduite	6	0,8
FA non utilisés comme des équivalents traductionnels	9	1,3
Total	718	100

Tableau 2 : Catégories de bruit relevées dans les FA extraits des sorties de AT1

Ces résultats indiquaient clairement qu'il fallait incorporer la catégorisation grammaticale. Nous avons donc remplacé l'analyseur de traductions AT1 par un nouvel analyseur, AT2, qui se distinguait du premier en ce que ses deux modèles de langage incorporaient le programme de catégorisation grammaticale de Foster [10]. On a ensuite modifié le processus de recherche pour tenir compte de l'information grammaticale associée à nos 145 paires de FA. Cette méthode a donné de bien meilleurs résultats, comme on peut le constater dans les Tableaux 3 et 4.

3. À noter, cependant, qu'aucun élément de «bruit» n'était attribuable à des alignements erronés, notre algorithme assurant un taux de réussite d'environ 98 p. cent.

	Nombre de cas	Pourcentage
Erreurs réelles	56	76,7
Bruit	17	23,3
Total	73	100

Tableau 3 : Résultats de l'extraction de FA dans les sorties de AT2

Catégorie de bruits	Nombre de cas	Pourcentage
Mauvaise catégorie grammaticale	6	35,3
Citation non traduite	2	11,8
FA non utilisés comme des équivalents traductionnels	9	52,9
Total	17	100

Tableau 4 : Catégories de bruit relevées dans les FA extraits des sorties de AT2

La catégorisation grammaticale a réduit considérablement le niveau de bruit, tout en n'exerçant qu'un effet marginal sur le nombre de cas rappelés (perte d'un cas). Cette amélioration spectaculaire est en grande partie attribuable à la résolution des problèmes liés à un petit nombre de mots utilisés fréquemment (comme le cas du mot *local*, mentionné précédemment). Une partie du bruit restant est due à des erreurs de catégorisation grammaticale, mais la majeure partie découle maintenant de la grossièreté de notre modèle de correspondance.

Il ne fait aucun doute que de meilleurs modèles permettraient d'améliorer la détection des faux amis. Cependant, la performance de la méthode de faible coût algorithmique que nous avons testée ici pourrait fort bien s'avérer suffisante pour les applications réelles.

5. La dictée de traductions : TransTalk

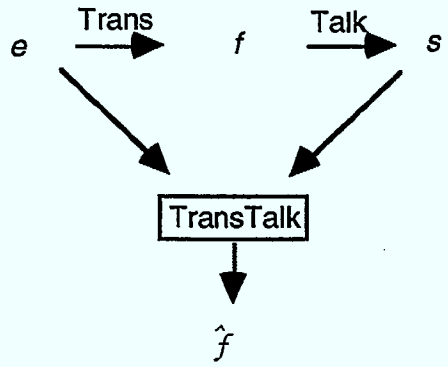
Nous sommes revenus quelques fois, dans cet article, sur le fait que des modèles de traduction faibles, s'ils sont employés de façon réaliste, peuvent offrir au traducteur des outils efficaces qui ne lui imposent pas de contraintes artificielles. Considérant que de nombreux traducteurs préfèrent

dicter leurs textes plutôt que de les dactylographier eux-mêmes, un module de dictée automatique constituerait une addition fort utile au poste de travail de traducteur (Gurstein et Monette [11]).

La technologie de la reconnaissance automatique de la parole est actuellement limitée aux vocabulaires restreints et elle est, de ce fait, inapplicable à la tâche de la plupart des traducteurs. Une possibilité intéressante, cependant, serait le *couplage* d'un module de reconnaissance de la parole et d'un modèle de traduction (faible). Le modèle de TA servirait alors à faire des prédictions probabilistes quant aux énonciations susceptibles d'être produites librement par le traducteur, afin de réduire dynamiquement le «vocabulaire réel probable» envisagé par le module de reconnaissance de la parole pour chaque unité de dictée (phrase ou paragraphe) et ce, jusqu'au point où la reconnaissance complète de ces unités pourrait être tentée.

Il est évident, par exemple, que la composition probabiliste du vocabulaire considéré par un module de reconnaissance de la parole qui tente de décoder la phrase *Ces impôts cachés doivent être acquittés par les pauvres aussi bien que par les riches* sera très différente selon que sa source anglaise *The poor as well as the rich have to pay these hidden taxes* est ou non connue du module. Il est beaucoup plus probable, par exemple, que la traduction française de cette phrase anglaise renferme le mot *impôts* qu'une phrase française choisie au hasard. Il semble donc raisonnable d'espérer qu'un modèle de traduction faible puisse rendre cette composition accessible au module de reconnaissance de la parole.
Cette idée a été avancée par Dymetman, Foster et Isabelle [9] ainsi que par Brown *et al.* [4]. Nous avons entrepris un projet de collaboration avec le groupe de reconnaissance de la parole du Centre de recherches informatiques de Montréal (CRIM), le projet *TransTalk*, qui vise à démontrer la faisabilité de cette approche, en utilisant l'anglais comme langue source et le français comme langue de dictée. Nous avons l'intention de nous limiter, au début, à la dictée de mots isolés, puis de passer progressivement à la parole continue. Les projets TransSearch et TransCheck décrits précédemment comportaient l'élaboration d'analyseurs de traductions incorporant des modèles de langage pour le français et l'anglais et un modèle de correspondance français-anglais (alignement de phrases) qui ont été entraînés sur le corpus du Journal des débats de la Chambre des communes. Ce corpus, ou domaine, est donc un choix tout naturel pour le projet TransTalk, puisque les modules existants constitueront alors des ressources fondamentales pour TransTalk. On peut en fait envisager TransTalk comme un système incorporant un analyseur de traductions fort semblable à ceux que nous avons déjà décrits, sauf qu'il a la capacité de traiter une langue cible parlée, plutôt qu'écrite.

TransTalk est basé sur un modèle probabiliste *p* de dictée de traduction qui met en relation une unité textuelle anglaise écrite *e*, sa traduction française écrite *f* (pour simplifier, nous supposerons que ces unités textuelles sont des phrases) et *s*, la réalisation acoustique de *f*. Les unités *e* et *s* sont toutes deux connues du système, qui a pour tâche de produire une estimation $\hat{f}$ de l'unité *f* effectivement formulée par le traducteur.



L'on est donc amené à définir $\hat{f}$ comme :

$$\hat{f} = \text{argmax}_f \; p(f \mid e, s)$$

c'est-à-dire que $\hat{f}$ est la phrase française la plus probable selon le modèle *p*, étant donné la phrase source anglaise et la réalisation acoustique de la phrase française.

En utilisant la formule de Bayes, on peut réécrire cette équation comme suit :

$$\begin{aligned} \hat{f} &= \text{argmax}_f \; p(s \mid e, f)\, p(f \mid e) \\ &= \text{argmax}_f \; p(s \mid f)\, p(f \mid e) \end{aligned}$$

la dernière égalité étant une conséquence de l'assomption modérée suivante : une fois que *f* est connue, les connaissances supplémentaires sur *e* n'ajoutent rien à la détermination de *s*.

Cette équation rappelle fortement la «formule fondamentale» de la reconnaissance statistique de la parole (Bahl *et al.* [2]) :

$$\hat{f} = \text{argmax}_f \; p(s \mid f)\, p(f)$$

où les distributions $p(s \mid f)$ et $p(f)$ sont appelées respectivement le modèle acoustique et le modèle de langage. Dans la situation envisagée ici, le modèle de langage pur $p(f)$ a été remplacé par un «modèle de langage conditionnel» $p(f \mid e)$, dans lequel la connaissance de e «affine» la structure statistique du modèle de langage, en le forçant en particulier à «concentrer» son attention sur un sous-ensemble lexical restreint de la langue. On peut mesurer quantitativement cet «affinement» au moyen de la *perplexité*, une quantité relative à la théorie de l'information qui mesure l'incertitude moyenne qu'un modèle de langage manifeste à l'égard du prochain mot devant apparaître dans un texte naturel, après avoir vu les mots précédents : moins un modèle est perplexe, plus il est prédictif (Jelinek [15]). Brown *et al.* [4] décrivent les résultats d'une expérience qu'ils ont réalisée sur le Journal des débats en utilisant un de leurs modèles de traduction les plus simples (du français àl'anglais, dans leur cas). Ces résultats révèlent que la perplexité par mot de leur modèle de langage pur (anglais) s'établit en moyenne à 63,3, tandis que la perplexité de leur modèle de langage conditionnel chute à une moyenne de 17,2. Ces résultats sont très encourageants pour la dictée, car ils signifient que le module acoustique devrait pouvoir faire un choix, étant donné un mot anglais parlé, parmi en moyenne 17;2 candidats équiprobables proposés par le modèle de langage conditionnel, par opposition à 63,3 candidats équiprobables proposés par le modèle de langage pur.

Plusieurs approches sont possibles pour la modélisation de $p(f \mid e)$. *Une première approche*, proposée par l'équipe IBM, consiste à utiliser la formule de Bayes pour écrire, par analogie à la formulation standard du problème de reconnaissance de la parole :

$$p(f \mid e) \propto p(e \mid f)\, p(f)$$

où (selon leur terminologie) $p(e \mid f)$ correspond au «modèle de traduction», qui joue un rôle semblable à celui du modèle acoustique dans la reconnaissance de la parole. Cela nous amène par conséquent à une formule symétrique pour l'ensemble du modèle de dictée de traduction, dans laquelle $p(f)$ est le modèle de langage, $p(s \mid f)$ est le modèle acoustique et $p(e \mid f)$ est le modèle de traduction. Cette méthode présente deux avantages : (1) elle repose sur un seul modèle de langage pour le français et (2) les travaux réalisés chez IBM sur la TA statistique semblent indiquer que des approximations même grossières de $p(e \mid f)$, *lorsqu'elles sont couplées à un bon modèle de langage pour le français*, donnent des approximations acceptables pour le modèle de langage conditionnel $p(f \mid e)$. C'est comme s'il y avait une «division du travail» entre $p(f)$, qui est responsable

de la structure correcte des phrases françaises, et $p(e \mid f)$, qui est chargé d'apparier les phrases anglaises et françaises (d'où le terme quelque peu trompeur de «modèle de traduction»), sans trop tenir compte de la structure interne du français ou de celle de l'anglais[4] (voir [9] pour plus de détails). Cette méthode présente toutefois une lacune importante au niveau du traitement : elle exige la réalisation d'une recherche étendue parmi les phrases f pour maximiser $p(e \mid f)\, p(f)$ (sans compter le facteur $p(s \mid f)$, ce qui ne fait qu'aggraver les choses). On sait que cela pose de sérieuses difficultés pratiques en termes de résultats de recherche non optimaux ainsi qu'en termes de temps de traitement, ce dernier facteur étant évidemment de toute première importance pour une application de dictée.

Une deuxième approche à la modélisation de $p(f \mid e)$ consiste à considérer *a priori* une certaine famille paramétrisée de modèles de langage $p_\lambda(f)$ pour le français, à décrire une mise en correspondance $e \rightarrow \lambda(e)$, puis à définir le modèle de langage conditionnel au moyen de :

$$p(f \mid e) = p_{\lambda(e)}(f)$$

Bien qu'elle présente l'inconvénient d'utiliser plus qu'un modèle de langage de référence pour le français, cette approche peut être mise en oeuvre efficacement si la famille $p_\lambda(f)$ est bien choisie. Nous examinons actuellement la possibilité d'adapter un modèle de langage proposé dans [10]. Ce modèle est une sorte de modèle markovien caché «tri-POS», qui dépend de deux familles de paramètres. La première famille $a_{i,\,j,\,k}$ donne la probabilité de générer un mot de catégorie grammaticale POS_k, les mots de catégories grammaticales POS_i et POS_j ayant déjà été générés. La deuxième famille, $b_{i,\,m}$, donne la probabilité qu'une catégorie grammaticale donnée POS_j soit associée au mot m. Cela signifie, sur le plan conceptuel du moins, que le modèle génère d'abord des chaînes de catégories grammaticales, en utilisant la fenêtre du contexte des deux catégories grammaticales déjà générées, puis «décore» chaque catégorie grammaticale avec un mot donné, dépendant uniquement de cette catégorie grammaticale. Les paramètres $a_{i,\,j,\,k}$ représentent une approximation de la structure «grammaticale» du français, tandis que les paramètres $b_{i,\,m}$ représentent une approximation de sa structure «lexicale».

---

4. En fait, il est facile de voir que, pour les fins de la traduction de l'anglais au français, il est équivalent d'utiliser $p(e \mid f)$ ou $p(e \mid f) \,/\, p(e)$ comme «modèle de traduction». Cette dernière quantité est l'exponentielle de l'information mutuelle entre e et f, une quantité symétrique dans e et f, qui ne possède aucune mémoire de la structure statistique interne de e ou de f, mais seulement de leur relation statistique.

Nous nous proposons de faire l'essai d'un schème où ces paramètres varieront dynamiquement selon la phrase source observée *e*. Une possibilité intéressante serait de maintenir les «paramètres grammaticaux» à leurs valeurs globales fixes en langue française (sans tenir compte de l'influence de la composition grammaticale de la phrase anglaise sur sa traduction), tout en modifiant les paramètres «lexicaux» selon la composition lexicale de la phrase anglaise. La première famille de paramètres peut être estimée de façon fiable sur un corpus français suffisamment étendu, tandis que la deuxième famille de paramètres, qui dépend de *e*, peut être estimée si l'on fait certaines hypothèses simplificatrices s'apparentant au modèle de traduction 1 de Brown *et al.* [5]. Essentiellement, chaque $b_{i,\ m\,f}\ (e)$ est considéré être la moyenne des contributions $p\ (m_f \mid m_e\ ,\ POS_i)$ faites par chaque mot anglais m_e de *e* à la probabilité de réaliser la catégorie grammaticale POS_i dans le mot français m_f. Pour estimer les paramètres $p\ (m_f \mid m_e\ ,\ POS_i)$, il faut partir d'un corpus d'apprentissage pré-aligné composé de bi-textes anglais-français (voir section 3). Il est alors possible de faire des estimations initiales pour les paramètres $p\ (m_f \mid m_e\ ,\ POS_i)$, puis d'utiliser des techniques de réestimation standard (voir [5]) sur ce corpus d'apprentissage pour maximiser l'efficacité prédictive de ces paramètres, tout en maintenant les paramètres grammaticaux à des valeurs fixes.

Le principal avantage de cette approche est que, pour chaque phrase source *e*, le modèle de langage conditionnel se réduit, en fait, à un simple modèle markovien caché $p_{\lambda(e)}(f)$; le problème de la dictée de traduction adopte alors une forme familière en reconnaissance de la parole, soit :

$$\hat{f} = argmax_f\, p_{\lambda(e)}(f)\ p\ (s \mid f)$$

pour laquelle il existe de puissantes techniques de recherche (Bahl *et al.* [2]).

6. Conclusions

Une nouvelle génération d'aides à la traduction se profile déjà à l'horizon. Grâce au développement des techniques d'analyse de traductions, les postes de travail de traducteurs pourront bientôt mettre à la disposition de leurs utilisateurs des outils qui dépasseront les simples fonctions de bureautique. Les traducteurs seront bientôt en mesure de tirer profit du vaste potentiel inexploité que recèle leur production antérieure. Ils disposeront bientôt d'outils de vérification qui les aideront à détecter les erreurs de traduction présentes dans leurs premiers jets. Et il y a de bonnes chances

que la transcription automatique de la parole se concrétise en traduction bien avant qu'elle ne devienne une réalité pour les applications unilingues.

Nous ne serions pas surpris de voir cette liste d'applications basées sur le concept de l'analyse de traductions s'allonger rapidement. Nous ne souhaitons que du bien à la TA classique, mais nous croyons que c'est dans le domaine des aides à la traduction que se produiront les véritables progrès, et ce, pendant de nombreuses années encore!

Références bibliographiques

[1] Bar-Hillel, Y. : «The State of Machine Translation in 1951», in *American Documentation*, vol. 2, 1951, p. 229-237.

[2] Bahl, L., Jelinek, F. et R. Mercer : «A maximum likelihood approach to continuous speech recognition», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2), p. 179-191, mars 1983.

[3] Brown, P., Lai, J. et R. Mercer : «Aligning sentences in parallel corpora», *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley (Californie), juin 1991.

[4] Brown, P., Chen, S., Della Pietra, S., Della Pietra, V., Kehler, S. et R. Mercer : «Automatic speech recognition in machine aided translation», 1992 (à paraître).

[5] Brown, P., Della Pietra, S., Della Pietra, V., et R. Mercer : «The Mathematics of Machine Translation: Parameter Estimation», (à paraître).

[6] Church, K. et W. Gale : «Concordances for Parallel Texts», in *Proceedings of the 7th Annual Conference of the UW Centre for NOED and Text Research*, Oxford, 1991.

[7] Debili, F. et E. Sammouda : «Appariement des phrases de textes bilingues français-anglais et français-arabes, in *Proceedings of COLING-92*, Nantes, 1992.

[8] Dymetman, M. : *Transformations de grammaires logiques et réversibilité en traduction automatique*, thèse d'État, Université de Grenoble 1, France, 1992.

[9] Dymetman, M., Foster, G. et P. Isabelle : *Towards an Automatic Dictation System for Translators (TransTalk)*, rapport technique, CITI (CITI), Laval (Québec), Canada, 1992.

[10] Foster, G. : *Statistical Lexical Disambiguation*, mémoire de maîtrise, McGill University, School of Computer Science, 1991.

[11] Gurstein, M. et M. Monette : *Functional Specifications for a Translator's Workstation*, rapport technique 12SD.36902-5-0003, Socioscope Inc., Ottawa, Canada, octobre 1988. Rapport présenté au Centre canadien de recherche sur l'informatisation du travail (Centre d'innovation en technologies de l'information).

[12] Isabelle, P. : «Machine Translation at the TAUM Group», in Margaret King (éd.), *Machine Translation Today: The State of the Art*, Edinburgh University Press, 1987.

[13] Isabelle, P. : «Bi-Textual Aids for Translators», in *Proceedings of the Eighth Annual Conference of the UW Centre for the New OED and Text Research*, University of Waterloo, Waterloo, Canada, 1992.

[14] Isabelle, P., Dymetman, M. et E. Macklovitch : «CRITTER: a Translation System for Agricultural Market Reports», in *Proceedings of COLING-88*, Budapest, 1988.

[15] Jelinek, F. : «Self-Organized Modeling for Speech Recognition», in Alex Waibel et Kai-Fu Lee, éditeurs, *Readings in Speech Recognition*, p. 450-506, Morgan Kaufman, San Mateo, Californie, 1990.

[16] Macklovitch, E. : «Corpus-Based Tools for Translators», in *Proceedings of the 33rd Annual Conference of the American Translators Association*, San Diego, 1992.

[17] Macklovitch, E. : *A Third Version of the CWARC's Workstation for Translators*, rapport technique, CWARC (CITI), Laval (Québec), Canada, 1993.

[18] Kay, M. : *The Proper Place of Men and Machines in Translation*, CSL-80-11, Xerox PARC, 1980.

[19] Sato, S. et M. Nagao : «Toward Memory-Based Translation», in *Proceedings of COLING-90*, p. 247-252, 1990.

[20] Simard, M., Foster, G. et P. Isabelle : «Using Cognates to Align Sentences in Parallel Corpora», in *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation*, Montréal, 1992.

[21] Van Noord, G. : *Reversibility in Natural Language Processing*, CIP-Gegevens Koninckliijke Bibliotheek, La Haye, 1993.

[22] Van Roey, J., Granger, S. et H. Swallow : *Dictionnaire des faux-amis français-anglais*, Paris, Duculot, 1988.

CACC / CCAC
35638

QUEEN P 98 .I814 1993 c.2
Isabelle, Pierre
L'analyse de traduction et l

DATE DUE

PRINTED IN U.S.A.

1575, boulevard Chomedey, Laval (Québec), H7V 2X2 Téléphone: (514) 973-5700 Télécopieur: (514) 973-5757

1575, Chomedey Boulevard, Laval, Quebec, H7V 2X2 Telephone: (514) 973-5700 Facsimile: (514) 973-5757