Meta
Journal des traducteurs
Translators' Journal

META

La génération de textes multilingues par un utilisateur monolingue

Harold Somers and Danny Jones

Volume 37, Number 4, décembre 1992

Études et recherches en traductique / Studies and Researches in Machine Translation

URI: https://id.erudit.org/iderudit/004200ar
DOI: https://doi.org/10.7202/004200ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)
1492-1421 (digital)

Explore this journal

Cite this article

Somers, H. & Jones, D. (1992). La génération de textes multilingues par un utilisateur monolingue. *Meta*, *37*(4), 647–656. https://doi.org/10.7202/004200ar

Article abstract

In this article we describe an approach to machine translation which involves multilingual text generation via interaction with a monolingual user: the system works in a specific and rather limited domain. Among the techniques used are the use of examples instead of linguistic rules to give the translation equivalents, and the encoding of contextual knowledge as a model of possible texts.

LA GÉNÉRATION DE TEXTES MULTILINGUES PAR UN UTILISATEUR MONOLINGUE

HAROLD SOMERS et DANNY JONES
Centre for Computational Linguistics, UMIST, Manchester, Angleterre

Résumé

Dans cet article, nous décrivons une approche de la traduction automatique qui comprend la génération de textes multilingues par une interaction avec un utilisateur monolingue : le système fonctionne dans un domaine spécifique et plutôt limité. Entre autres grandes techniques employées, on trouve l'utilisation d'exemples au lieu de règles linguistiques pour donner les équivalents entre les langues, et le codage des connaissances contextuelles comme un modèle de textes possibles.

Abstract

In this article we describe an approach to machine translation which involves multilingual text generation via interaction with a monolingual user: the system works in a specific and rather limited domain. Among the techniques used are the use of examples instead of linguistic rules to give the translation equivalents, and the encoding of contextual knowledge as a model of possible texts.

1. AVANT-PROPOS

Cet article décrit les grandes lignes d'un système[1] qui contient plusieurs nouvelles approches de la traduction automatique (TA) : c'est un système de génération de textes multilingues pour un utilisateur monolingue qui fonctionne dans un domaine limité et qui permet à l'utilisateur de composer, à l'aide d'un dialogue, des textes de bonne tenue et bien mis en page, en plusieurs langues, y compris celle de l'utilisateur lui-même. La spécificité du domaine permet au système d'interpréter les entrées de l'utilisateur, et même de prédire le contenu du texte, deux traits qui contribuent à obtenir un niveau de traduction élevé.

Le domaine qui nous concerne actuellement est la génération des petites annonces d'emploi dans des secteurs techniques comme l'informatique, l'administration ou les arts graphiques, où la nouvelle flexibilité professionnelle qu'offre la situation imminente en Europe a suggéré une application évidente. Un avantage de ce domaine, qu'ont étudié également Shann et Warwick (1983), est que les textes sont de forme assez fixe mais dans une certaine mesure de contenu varié, combinaison qui convient très bien à notre approche. D'autres domaines ont des propriétés similaires, comme les suivants :

- l'échange de messages dans une compagnie multinationale, par exemple entre les fournisseurs pour la disponibilité et les commandes de marchandises, de pièces détachées, etc., ou pour des offres de services, d'équipement, etc. ;
- les premiers rapports d'entretien ou de panne des clients ou des techniciens de maintenance (on pense aux grands fabricants d'appareils électroniques basés dans un pays anglophone qui exportent partout dans le monde) ;

- la communication dans les services militaires ou policiers (par exemple l'OTAN, l'Interpol, ou la nécessité de la communication anglais-français dans le Tunnel sous la Manche).

Le mode d'opération est celui d'un système interactif : une interface flexible, pilotée soit par l'utilisateur soit par le système, avec laquelle ils coopèrent afin de garantir la qualité de la traduction. Les bases de connaissances linguistiques et spécifiques au domaine sont fortement liées lorsque les équivalents bilingues sont liés aux contextes de situation. De plus, le système se sert des connaissances du domaine pour indiquer à l'utilisateur les formes acceptables et standardisées. Le système est néanmoins assez flexible : il ne s'agit pas de produire des messages rigides avec des phrases préprogrammées ! Le système accepte l'anglais, et génère des textes en anglais, en espagnol, et en grec. Le traitement du français est aussi prévu.

2. CONTEXTE

La conception du système répond aux faiblesses perçues dans les approches classiques à la TA, que l'on trouve encore dans beaucoup de systèmes actuels. Bien que l'on admette depuis longtemps que la traduction de haute qualité entièrement automatique — la célèbre FAHQT (*«fully automatic high quality translation»*) — soit irréalisable, les solutions proposées — dont les plus connues sont l'interaction avec l'utilisateur et les restrictions sur les textes d'entrée — et les conceptions linguistiques et informatiques de la «deuxième génération» (Vauquois 1975) nous semblent maintenant être arrivées à leurs limites.

La solution interactive se présente en général sous la forme de la boîte à outils du traducteur (*«translator's workbench»*) (Kay 1980, Melby 1982) dont le but est d'aider ce dernier à traduire. Souvent, un tel système propose sa traduction suivie d'une phase plus ou moins conviviale de postédition ; dans les cas les plus ambitieux, l'interaction entre le système et l'utilisateur a lieu au cours du traitement lui-même, où l'utilisateur répond aux questions concernant la désambiguïsation du texte source et les problèmes lexicaux et stylistiques du texte cible. Le système et l'utilisateur doivent l'un et l'autre connaître les deux langues, et il est souvent difficile d'assurer une division du travail correcte. Les connaissances de l'utilisateur et celles du système peuvent même être en conflit. L'expérience a démontré que les utilisateurs — même ceux qui étaient au début très en faveur — finissent par devenir frustrés par un système qui n'apprend jamais rien et recommence les mêmes erreurs banales.

La TA de textes limités (TATL) a connu sa plus grande réussite avec le système MÉTÉO (Chandioux et Guérard 1981), qui a promu la notion de «sous-langues» (Kittredge 1987), ce qui a suscité l'intérêt de plusieurs chercheurs (Isabelle 1989, Kosaka *et al.* 1988 ; voir Luckhardt 1991). Ce sont les systèmes où la couverture syntaxique et lexicale est décidée par les caractéristiques de textes typiques d'un certain domaine. Cependant, pour la plupart des systèmes de TATL, les limitations sont dictées par les lacunes des systèmes plutôt que par le contenu des textes à traiter (voir Elliston 1979, Pym 1990, Sager 1990 : 7) (honorable exception : TITUS — Ducrot 1974, 1988).

On peut aussi critiquer les systèmes de TA de deuxième génération en ce qui concerne l'informatique et la linguistique : les méthodes typiques reflètent les techniques préférées des années 60 et 70, maintenant dépassées, et qui envisagent un traitement où, par défaut, on préserve la structure originale dans le texte cible, c'est-à-dire que nous «produisons une traduction la plus littérale possible» (*cf.* Somers 1987 : 84).

3. NOUVELLES IDÉES CONTRIBUTIVES

3.1. TA POUR MONOLINGUES

Au lieu de la boîte à outils du traducteur, on a commencé, il y a sept ans à peu près, à concevoir des systèmes interactifs pour des utilisateurs monolingues qui traduisent leurs propres textes; une idée lancée plusieurs années plus tôt par Martin Kay (1973), et puis adoptée vers 1985 simultanément par un bon nombre de chercheurs (y compris Carbonell et Tomita 1987, Johnson et Whitelock 1987, Schubert 1986, Whitelock *et al.* 1986, Zajac 1986). Dans un tel système, on suppose que l'utilisateur sait bien ce qu'il veut dire mais qu'il ne sait pas comment le dire dans la langue cible, alors que le système sait bien comment traduire, mais demande à l'utilisateur de l'aider à comprendre le texte source. Dans ce scénario, il est évident que l'interaction homme-machine risque d'être assez complexe, ce qui a donné naissance à l'idée de la TA basée sur des dialogues (TABD) (Blanc et Boitet 1990, Boitet 1990). Dans un cas extrême (Somers *et al.* 1990), le système et l'utilisateur travaillent en collaboration pour composer le texte source en même temps que pour le traduire.

3.2. GÉNÉRATION DE TEXTES MULTILINGUES

S'il n'y a pas de texte source, l'attention se concentre sur la génération (ou synthèse) du ou des texte(s) cible(s), partie de la TA qui, comparée à l'analyse et au transfert, a été méconnue (*cf.* Macdonald 1987). Un bon nombre de chercheurs ont eu l'idée de concevoir des systèmes où des textes assez stéréotypés dans des domaines fixes peuvent être générés à l'aide d'une interaction avec un utilisateur. Ces systèmes sont, pour la plupart, pilotés par menu: c'est le cas de notre propre système qui génère de la correspondance commerciale en français, en allemand et en espagnol (Jones et Tsujii 1990), de celui de Saito et Tomita (1986), qui produit des textes en japonais, et du système de Zaki et Noor (1991), pour composer des lettres officielles en malais. Dans notre système, des textes originaux peuvent être construits par l'assemblage d'extraits de textes déjà en mémoire. Le grand avantage de cette variation de la TA, c'est que les textes sont de très bonnes «traductions», étant donné qu'elles sont copiées d'un corpus de vraies traductions (c'est-à-dire faites par des humains), plutôt que construites selon des règles abstraites conçues par des linguistes. Ces systèmes représentent alors une forme primitive de...

3.3. TA BASÉE SUR DES EXEMPLES (TABE)

L'aspect le plus important du système que nous présentons actuellement est qu'il s'agit d'un système de TABE. L'idée originale est de Nagao (1984) et se trouve depuis deux ou trois années seulement dans les programmes de recherches de Jones (1991), Kitano et Higuchi (1991), Sadler (1989, 1991), Sato et Nagao (1990), Sumita *et al.* (1990) et Sumita et Iida (1991). Il s'agit toujours d'un corpus bilingue de textes déjà traduits par des humains, mais, au lieu d'en tirer de simples modèles, on l'utilise comme base de connaissances linguistiques pour suggérer une traduction correcte. Il se peut, par exemple, que le corpus fournisse un exemple sémantique assez similaire à ce qu'on veut traduire: c'est le cas de Sadler et de Sumita; ou bien il s'agit de recombiner des fragments grammaticaux, comme le font Jones et Sato/Nagao.

Dans la TABE, les problèmes les plus importants sont les suivants: comment retrouver d'une manière efficace les exemples assez similaires au texte donné, puis comment être sûrs de la façon de les recoller. Le premier problème est très bien connu dans le cadre de la TA basée sur corpus, dont la TABE est un cas spécial. Brown *et al.* (1990) ont travaillé chez IBM avec un corpus parallèle de 100 millions de mots en anglais et en français tiré du *Hansard* canadien, c'est-à-dire des actes du Parlement à Ottawa. Ils ont voulu faire de la TA par des méthodes purement statistiques, ce qui les a obligés à trouver

des moyens d'aligner automatiquement le corpus bilingue (Brown *et al.* 1991), problème qu'ont signalé également Catizone *et al.* (1989), Chen *et al.* (1991), Gale et Church (1991) et Kay et Röscheisen (1988).

On a déjà noté l'avantage de la TABE où les résultats sont tirés des vraies traductions, ce qui veut dire que le style est garanti et que la traduction ne se contente pas de préserver la structure du texte source, comme dans la TA traditionnelle. De plus, pour augmenter un système de TABE, il suffit d'ajouter de nouveaux exemples dans la base de données. On évite le problème d'entropie de performance d'un système traditionnel, où ajouter de nouvelles règles provoque de nouvelles erreurs, qui peuvent ne se présenter que beaucoup plus tard, ce qui rend le débogage très difficile.

4. DESCRIPTION DU SYSTÈME

Supposons d'abord que les utilisateurs du système ne parlent qu'une langue, contrairement aux traducteurs. Une remarque immédiate : nous n'avons plus la possibilité de consulter l'utilisateur pour demander si la traduction est correcte. Comment assurer la traduction dans ce cas ? Les moyens les plus naturels sont les exemples, le dialogue entre le système et l'utilisateur, et l'exploitation des connaissances du contexte et du domaine.

L'utilisation des exemples de traductions a comme avantage l'exploitation maximale des régularités superficielles de la sous-langue donnée. Ces exemples fournissent non seulement les constructions grammaticalement correctes dans la langue cible, mais aussi les facteurs de style propres au type de texte. L'utilisation des exemples pose cependant deux grands problèmes : premièrement, comment les associer d'une façon souple avec les entrées, et leur couverture.

On peut reformuler le premier problème de la façon suivante : «Quels sont les exemples dans la base de données qui s'associent avec cette entrée ?» Étudions pour commencer la méthode de représentation des exemples. Dans notre système, elle a lieu à des niveaux syntaxiques variés qui diffèrent selon leur degré d'abstraction, par exemple une simple chaîne de caractères, un cadre syntaxique avec des choix pour des alternatives restreintes, ou bien une formule représentant les meilleures formes superficielles du cadre sémantique de l'exemple (avec les équivalences dans la langue cible). Avec une telle gamme de représentations, l'appariement sera possible dans la grande majorité des cas, bien que le nombre de candidats augmente à mesure que s'accroît le niveau d'abstraction.

Les connaissances du contexte sont utilisées pour réduire l'espace de recherche. À côté de la description syntaxique, chaque exemple est associé avec une représentation qui indique ses relations avec le texte dont il est originaire. Afin d'utiliser cette information, un «modèle d'intention» permet au système de comprendre la nature du texte qu'il essaie de composer (et éventuellement de traduire). Par exemple, certaines relations contextuelles se présentent simultanément dans certaines circonstances (à l'intérieur de la sous-langue) : une fois l'exemple avec son indicateur de contexte trouvé, le modèle d'intention permettra au système de s'attendre à certaines relations contextuelles plutôt qu'à d'autres, ce qui réduit d'autant le traitement de l'entrée suivante. Les connaissances du contexte déterminent également le choix correct dans les cas où une entrée peut correspondre à plusieurs candidats.

Le second problème se formule ainsi : «Est-on sûr de toujours trouver dans la base de données des exemples qui s'associent avec telle entrée ?», et c'est le problème fondamental de la TABE lorsque, à l'extrême, elle ne tient aucun compte des généralités linguistiques. On peut le surmonter en appuyant le processus de composition des textes source et cible sur la «*recombination*» (*cf.* Jones 1991). Ce but est réalisé à l'aide des représentations des exemples, des définitions des relations contextuelles, et du modèle

global d'intention. Il est aussi important de s'assurer de la légalité grammaticale, pragmatique et stylistique de la recombinaison des exemples, et de maintenir en même temps la cohésion du texte. Dans certains cas, il faudra que l'utilisateur s'exprime différemment afin de se rapprocher de l'entrée attendue.

Pour recombiner ou changer l'entrée pendant le processus, il faut consulter l'utilisateur. Le système doit donc savoir interagir de façon intelligente : c'est le rôle du module additionnel qu'est le modèle de dialogue homme-machine. Pour que l'interaction soit «intelligente», le système doit comprendre l'intention de l'utilisateur, ce qui nécessite des connaissances du domaine.

Un point constant de cette recherche a été l'exigence de l'approche sous-langue. C'est maintenant un des aspects les moins contestés des recherches en TA, mais nous insistons sur le fait que la sous-langue n'est pas un simple prétexte pour limiter le vocabulaire et la couverture syntaxique du système. Nous la considérons plutôt comme le facteur essentiel qui lie les connaissances contextuelles exprimées dans le modèle d'intention et les connaissances linguistiques des représentations, tout en se rapportant aux connaissances générales du domaine. Le système est transportable dans la mesure où la sous-langue et ses ramifications peuvent être paramétrisées.

La conception générale du système est présentée dans la Figure 1 de l'annexe.

Le système a deux grandes composantes : le générateur de textes multilingues et le moniteur de dialogue, qui interagissent et partagent plusieurs sources de connaissances. L'emploi du système dépend quelque peu de la compétence de l'utilisateur, et nous pouvons distinguer deux scénarios extrêmes, avec des variations entre eux.

Dans le premier, l'utilisateur expérimenté propose son brouillon. L'algorithme d'appariement essaie de trouver dans le corpus d'exemples un ensemble d'exemples, dans la langue de l'utilisateur, qui correspondent bien avec l'entrée proposée. Les exemples sont stockés comme des paires de descriptions syntaxiques et pragmatiques de textes : les «descriptions syntaxiques» s'étendent sur des fragments entiers de textes, par des modèles partiels, jusqu'à des représentations linguistiques plus abstraites. Les «descriptions pragmatiques» associent l'information pragmatique ou contextuelle aux descriptions linguistiques, lesquelles servent à repérer les fragments de texte dans le «modèle d'intention» plus abstrait, qui détermine les structures de texte possibles en général. Le modèle d'intention sert donc à restreindre l'espace de recherche des exemples envisagé par l'algorithme d'appariement. Étant donné que les corpus ne sont pas composés de textes en parallèle, mais de collections de textes qui se ressemblent pragmatiquement, c'est davantage le modèle d'intention qui sert à donner les liens bi- et multilingues, bien qu'il y ait aussi quelques paires d'exemples — par exemple, des phrases fixes et des unités lexicales individuelles — qui fournissent, le cas échéant, des «traductions» au sens conventionnel.

Les résultats de la phase d'appariement sont passés au moniteur de dialogue, qui vérifie la proximité de l'entrée de l'utilisateur avec des entrées acceptables, étant donné la situation actuelle du texte : les connaissances du moniteur ont deux sources, à savoir le modèle d'intention, et les connaissances du domaine. Le moniteur de dialogue peut interagir avec l'utilisateur afin de confirmer l'entrée, ou de la rendre plus conforme aux attentes. Le texte «source» accepté, il reste la génération des textes cibles qui lui correspondent. Cette tâche entraîne de trouver des fragments d'exemples correspondants dans le corpus en langue cible et — toujours parce que les corpus ne sont pas parallèles — de les recombiner pour donner des textes cibles appropriés.

Dans le second scénario, l'utilisateur est moins expérimenté, et c'est le système qui prend l'initiative. Dans ce cas, le modèle d'intention et les connaissances du domaine pris ensemble fournissent une base de données qui munit le système d'un modèle prédictif du

texte qu'il utilise pour souffler à l'utilisateur d'une façon plus ou moins explicite les entrées propres. Dans le cas le plus extrême, le système peut dériver lui-même une série de menus ou d'hypertextes à lui présenter. Sinon, le système souffle des suggestions plus générales qui esquissent la fonctionnalité de chaque portion de texte, en laissant à l'utilisateur le soin de proposer son brouillon pour le segment. Le rôle du moniteur de dialogue et la génération des textes cibles sont les mêmes qu'auparavant.

5. DISCUSSION

Le système exploite une variété de techniques, dont beaucoup sont au premier plan des recherches. Pour la tâche de l'appariement de l'entrée de l'utilisateur avec les exemples stockés, on pourrait employer l'analyse grammaticale traditionnelle, mais la technique principale entraînera des méthodes stochastiques de l'appariement de modèles, etc. Le but est d'associer les entrées avec des exemples similaires, mais pas forcément identiques, dans la base de données; les techniques d'appariement doivent donc être assez flexibles pour trouver un choix de candidats pour une entrée donnée. C'est pourquoi une mesure de similarité est indiquée, selon les techniques proposées par Carroll (1990), Kay et Röscheisen (1988), Kitano et Higuchi (1991) et Skousen (1989) nous expérimentons également une approche connexioniste de ce problème (McLean 1992).

Les données du corpus multilingue dérivent d'une façon empirique des spécimens d'annonces d'emploi réelles. Pour des raisons pratiques, nous ne pouvons pas espérer trouver un corpus vraiment parallèle dans ce domaine. Les connaissances linguistiques comparatives du système ne peuvent donc pas se trouver dans des paires d'équivalents traductionnels, comme dans le système d'IBM (Brown *et al.* 1990, 1991), basé sur des méthodes statistiques. On compte alors sur le modèle abstrait d'intention comme sur une sorte de médiateur, et c'est la propriété fonctionnelle plutôt que formelle du fragment de texte qui donne son équivalent dans la langue cible. L'analyse des corpus multilingues que font Alexa et Bárcena (1992) fournit des données qui permettent la conception des représentations linguistiques pour les exemples, détermine le contenu et la forme du ou des modèle(s) d'intention — où les aspects fonctionnels et pragmatiques des annonces sont définis — et nous munit des informations nécessaires pour définir les connaissances du domaine. Celles-ci sont basées sur le contenu propositionnel du corpus, qui détermine non seulement ce que les langues du domaine ont en commun, mais aussi les absences d'équivalence, par exemple en ce qui concerne les titres des travaux et les qualifications. Un autre centre d'intérêt qui ressort du domaine choisi est la nécessité de refléter dans les modèles d'intention les différences culturelles que l'on trouve dans les annonces. Il se peut qu'elles soient superficielles, comme l'ordre typique des informations, mais on peut également prévoir des difficultés plus importantes (voir ci-dessous).

À propos de la génération des nouveaux exemples par «recombinaison», il ne s'agit pas ici de la génération d'après les représentations, lesquelles sont données a priori dans le corpus, mais de la capacité d'un tel mécanisme de générer des textes qui ne se trouvent pas directement dans l'ensemble d'exemples. On a déjà mentionné les avantages généraux du traitement du langage naturel basé sur des exemples. Il est reconnu cependant que, pour être vraiment flexibles, ces systèmes nécessitent une certaine capacité de génération au delà de celle fournie par les exemples «statiques». Cette capacité s'accroît par la combinaison des composants de plusieurs exemples. Cela arrive lorsque le texte d'entrée ne correspond pas à un seul exemple entier, mais que plusieurs exemples sont associés avec des parties de l'entrée. Évidemment, il faut ne pas rejeter le texte comme étant «mal formé», mais essayer de générer un «clone» (*cf.* Jones 1991) de l'entrée basée sur les meilleurs exemples trouvés pendant le processus d'appariement. Ce processus est guidé par le modèle d'intention et par les connaissances du domaine.

Celles-ci comprennent les connaissances linguistiques et pragmatiques comprises dans les modèles d'intention, mais aussi les connaissances plus générales concernant la rédaction d'une annonce d'emploi, nécessaires au moniteur de dialogue pour interpréter l'entrée de l'utilisateur et ses réponses durant les interactions, et pour construire un modèle des croyances de l'utilisateur, étant donné que c'est un cas typique de dialogue orienté vers une tâche. En sus — ce qui est surtout intéressant —, les connaissances du domaine doivent embrasser les connaissances culturelles comparatives concernant les différences de conventions dans les annonces d'emploi et les problèmes qui s'y rattachent, comme l'absence d'équivalences de qualifications entre les différentes communautés, ou bien les différences culturelles relatives au contenu du texte lui-même : il peut être acceptable de stipuler certaines conditions dans un pays alors qu'il est illégal de le faire dans un autre (par exemple, en espagnol les termes *señora*, *señorita*, *chica* et *muchacha*, qui se trouvent dans les annonces, ont des implications précises pour la candidate éventuelle — son âge, son sexe et même son état civil — qu'on n'a pas le droit d'exprimer dans une annonce similaire en Grande-Bretagne). Ces problèmes tout à fait pratiques doivent être résolus par les connaissances du domaine que le moniteur de dialogue va utiliser pour demander à l'utilisateur, le cas échéant, de s'exprimer différemment.

Ce projet de recherches comporte plusieurs autres aspects intéressants que nous ne pouvons pas considérer ici : notamment le côté dialogue du système. Le projet nous intéresse parce qu'il permet d'explorer quelques idées nouvelles, mais aussi parce que nous avons l'intention d'implémenter une maquette qui va démontrer une approche à la TA assez différente des autres approches courantes : la TA vue comme génération de textes multilingues.

Note

1. Une partie de ce travail est financée par le *Science and Engineering Research Council* du Royaume-Uni. Nous tenons à remercier les collègues suivant qui ont participé aux discussions où ce système a été conçu : Joseba Abaitua, Melina Alexa, Elena Bárcena, Nuira Bel, John Darzentas, Anne de Roeck, Rod Johnson, Ian McLean, et Mike Rosner. Nous remercions également Charles Boisvert, qui nous a aidé à rédiger ce texte en français. Toute faute qui reste, ainsi que les opinions exprimées, n'engagent que les auteurs.

BIBLIOGRAPHIE

ALEXA, Melina et Elena BÁRCENA (1992) : «Sublanguage and Multilingual Corpus Analysis for Example-Based Machine Translation», Rapport n° 92 / 3, Centre for Computational Linguistics, UMIST, Manchester.

BLANC, E. et C. BOITET (Eds) (1990) : *DBMT-90: Post-COLING Seminar on Dialogue-Based MT (Machine Translation of / with Dialogues)*, Grenoble, Groupe d'étude pour la traduction automatique.

BOITET, Christian (1990) : «Towards Personal MT: General Design, Dialogue Structure, Potential Role of Speech», *COLING-90*, Helsinki, vol. 2, pp. 30-35.

BROWN, Peter, COCKE, John, DELLA PIETRA, Stephen A., DELLA PIETRA, Vincent J., JELINEK, Frederick, LAFFERTY, John D., MERCER, Robert L. et Paul S. ROOSSIN (1990) : «A Statistical Approach to Machine Translation», *Computational Linguistics*, 16, pp. 79-85.

BROWN, Peter F., LAI, Jennifer C. et Robert L. MERCER (1991) : «Aligning Sentences in Parallel Corpora», *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, Californie, pp. 169-176.

CARBONELL, Jaime G. et Masaru TOMITA (1987) : «Knowledge-Based Machine Translation, the CMU Approach», Sergei Nirenburg (Ed.), *Machine Translation: Theoretical and Methodological Issues*, Cambridge, Cambridge University Press, pp. 68-89.

CARROLL, Jeremy J. (1990) : «Repetitions Processing Using a Metric Space and the Angle of Similarity», Rapport n° 90 / 3, Centre for Computational Linguistics, UMIST, Manchester.

CATIZONE, Roberta, RUSSELL, Graham et Susan WARWICK (1989) : «Deriving Translation Data from Bilingual Texts», *Proceedings of the First International Acquisition Workshop*, Detroit.

CHANDIOUX, John et Marie-France GUÉRAUD (1981) : «Météo : un système à l'épreuve du temps», *Meta*, 26-1, Presses de l'Université de Montréal, pp. 18-22.

CHEN, Shu-Chuan, CHANG, Jing-Shin, WANG, Jong-Nae et Keh-Yih SU (1991) : «ArchTran: a Corpus-Based Statistics-Oriented English-Chinese Machine Translation System», *Proceedings of the Machine Translation Summit III*, Washington DC, pp. 33-40.

DUCROT, Jean-Marie (1974) : «Perspectives et avantages offerts par la traduction automatique des analyses de documents selon la méthode TITUS», *Actes du 1er Congrès national français sur l'information et la documentation*, Paris, pp. 321-333.

DUCROT, Jean-Marie (1988) : «Le système TITUS IV : Système de traduction automatique et simultanée en quatre langues», *La traduction assistée par ordinateur : perspectives technologiques, industrielles et économiques envisageables à l'horizon 1990* (Actes du Séminaire international, Paris, et dossiers complémentaires), Paris, DAICADIF, pp. 55-67.

ELLISTON, John S.G. (1979) : «Computer-Aided Translation: a Business Viewpoint», Barbara M. Snell (Ed.), *Translating and the Computer*, Amsterdam, pp. 149-158.

GALE, William A. et Kenneth W. CHURCH (1991) : «A Program for Aligning Sentences in Bilingual Corpora», *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, Californie, pp. 177-184.

ISABELLE, Pierre, COCHARD, Jean-Luc, DYMETMAN, Marc, MACKLOVITCH, Elliott, PERRAULT, François et Michel SIMARD (1989) : «CRITTER : un système de traduction pour les rapports de marchés agricoles», *Actes du Congrès canadien en génie électrique et informatique*, Montréal, pp. 109-113.

JOHNSON, Roderick L. et Peter WHITELOCK (1987) : «Machine Translation as an Expert Task», Sergei Nirenburg (Ed.), *Machine Translation: Theoretical and Methodological Issues*, Cambridge, Cambridge University Press, pp. 136-144.

JONES, D. (1991) : *The Processing of Natural Language by Analogy with Specific Reference to Machine Translation*, Thèse de doctorat, UMIST, Manchester.

JONES, D. et J. TSUJII (1990) : «Interactive High Quality Translation for Monolinguals», *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Austin, Texas, pp. 43-46.

KAY, Martin (1973) : «The MIND System», Randall Rustin (Ed.), *Natural Language Processing*, New York, Algorithmics Press, pp. 155-188.

KAY, Martin (1980) : «The Proper Place of Men and Machines in Language Translation», Rapport de Recherche n° CSL-80-11, Xerox Palo Alto Research Center, Palo Alto, Californie.

KAY, Martin et Martin RÖSCHEISEN (1988) : «Text-Translation Alignment», Mémo de recherche, Xerox Palo Alto Research Center, Palo Alto, Californie.

KITANO, H. et T. HIGUCHI (1991) : «Massively Parallel Memory-Based Parsing», *Proceedings of the International Joint Conference on Artificial Intelligence IJCAI-91*, Sydney.

KITTREDGE, Richard I. (1987) : «The Significance of Sublanguage for Automatic Translation», Sergei Nirenburg (Ed.), *Machine Translation: Theoretical and Methodological Issues*, Cambridge, Cambridge University Press, pp. 59-67.

KOSAKA, Michiko, TELLER, Virginia et Ralph GRISHMAN (1988) : «A Sublanguage Approach to Japanese-English Machine Translation», Dan Maxwell, Klaus Schubert et Toon Witkam (Eds.), *New Directions in Machine Translation*, Dordrecht, Foris, pp. 109-120.

LUCKHARDT, Heinz-Dirk (1991) : «Sublanguages in Machine Translation», *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*, Berlin, pp. 306-308.

MACDONALD, David D. (1987) : «Natural Language Generation: Complexities and Techniques», Sergei Nirenburg (Ed.), *Machine Translation: Theoretical and Methodological Issues*, Cambridge, Cambridge University Press, pp. 192-224.

McLEAN, Ian J. (1992) : «Example-Based Machine Translation Using Connectionist Matching», *Actes du Quatrième Colloque international sur les aspects théoriques et méthodologiques de la traduction automatique : Méthodes empiricistes versus méthodes rationalistes en TA*, Montréal, pp. 35-43.

MELBY, Alan K. (1982) : «Multi-Level Translation Aids in a Distributed System», *Coling '82*, Prague, pp. 215-220.

NAGAO, Makoto (1984) : «A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle», A. Elithorn (Ed.), *Artificial and Human Intelligence*, Amsterdam, Elsevier, pp. 173-180.

PYM, P.J. (1990) : «Pre-Editing and the Use of Simplified Writing for MT: an Engineer's Experience of Operating an MT System», Pamela Mayorcas (Ed.), *Translating and the Computer 10: The Translation Environment Ten Years On*, Londres, Aslib, pp. 80-96.

SADLER, Victor (1989) : *Working with Analogical Semantics: Disambiguation Techniques in DLT*, Dordrecht, Pays Bas, Foris.

SADLER, Victor (1991) : «The Textual Knowledge Bank: Design, Construction, Applications», *Proceedings of the International Workshop on Fundamental Research for the Future Generation of Natural Language Processing*, Kyoto, Japon, pp. 17-32.

SAGER, Juan (1990) : «Ten Years of Machine Translation Design and Application: From FAHQT to Realism», Pamela Mayorcas (Ed.), *Translating and the Computer 10: The Translation Environment Ten Years On*, Londres, Aslib, pp. 3-10.

SAITO, H. et M. TOMITA (1986) : «On Automatic Composition of Stereotypic Documents in Foreign Languages» (Communication présentée à la *1st International Conference on Applications of Artificial Intelligence to Engineering Problems*, Southampon), Rapport de Recherche n° MU-CS-86-107, Pittsburg, Carnegie Mellon University.

SATO, Satoshi et Makoto NAGAO (1990) : «Toward Memory-Based Translation», *COLING-90*, Helsinki, vol. 3, pp. 247-252.

SCHUBERT, Klaus (1986) : «Linguistic and Extra-Linguistic Knowledge», *Computers and Translation*, 1, pp. 125-152.

SHANN, Patrick et Susan WARWICK (1983) : «Études contrastives dans un système de traduction automatique (SEPPLI)», Actes du Colloque «*Linguistique Contrastive et Coopération Culturelle*», Paris, *Contrastes*, Hors Série A3, pp. 61-72.

SKOUSEN, Royal (1989) : *Analogical Modeling of Language*, Dordrecht, Kluwer.

SOMERS, Harold (1987) : «Some Thoughts on Interface Structure(s)», Wolfram Wilss et Karl-Dirk Schmitz (Eds.), *Maschinelle Übersetzung — Methoden und Werkzeuge*, Tübingen, Niemeyer, pp. 81-99.

SOMERS, Harold L., TSUJII, Jun-ichi et Danny JONES (1990) : «Machine Translation without a Source Text», *COLING-90*, Helsinki, vol. 3, pp. 271-276.

SUMITA, Eiichiro et Hitoshi IIDA (1991) : «Experiments and Prospects of Example-Based Machine Translation», *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, Californie, pp. 185-192.

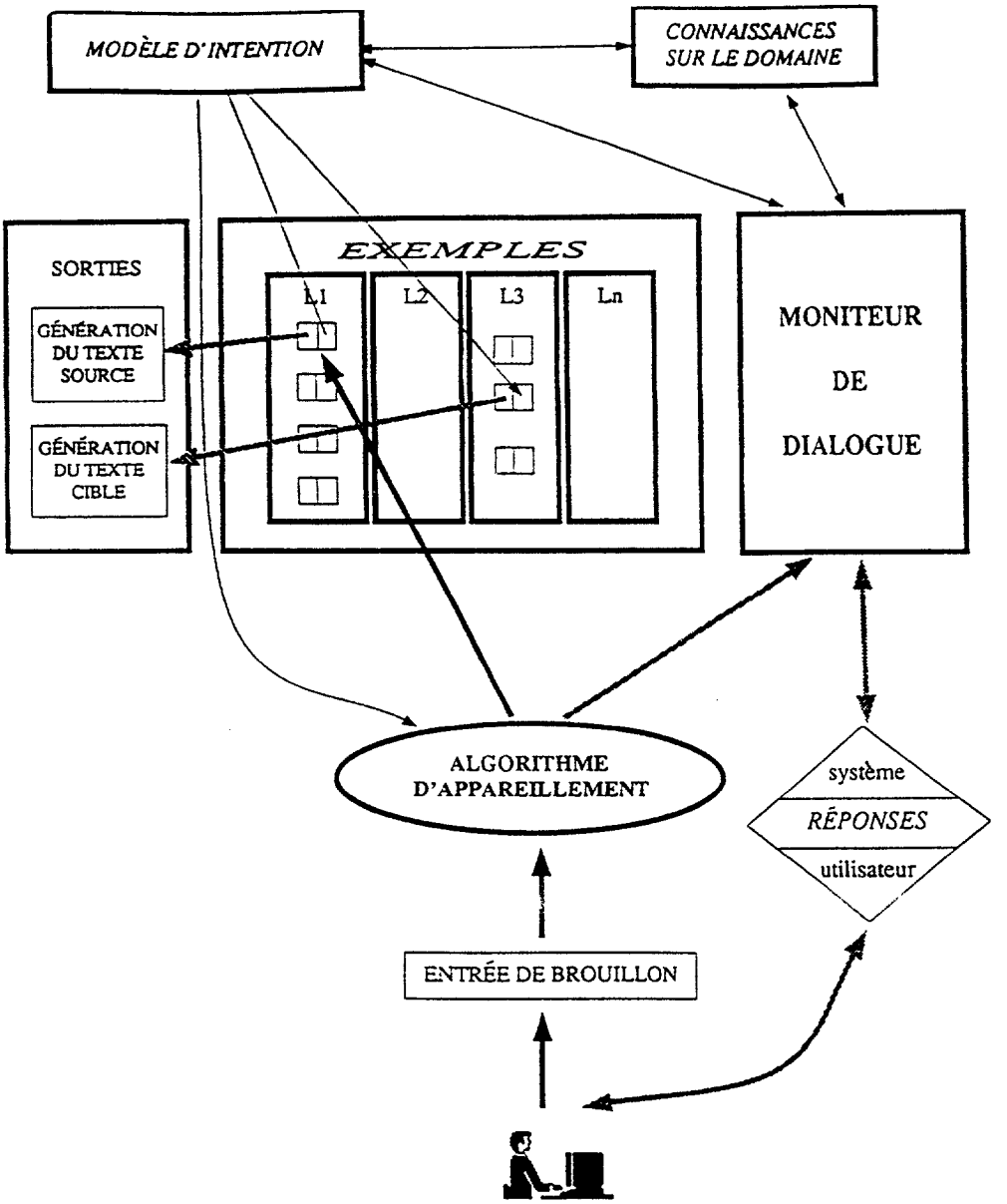
SUMITA, Eiichiro, Hitoshi IIDA et Hideo KOHYAMA (1990) : «Translating with Examples: a New Approach to Machine Translation», *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Austin, Texas, pp. 203-212.

VAUQUOIS, Bernard (1975) : «Méthodes, résultats et bilan de la traduction automatique», *Journées d'études du Festival international du son haute-fidélité, stéréophonie*, Paris, repris dans C. Boitet (Ed.), *Bernard Vauquois et la TAO : Vingt-Cinq Ans de Traduction Automatique — Analectes*, Grenoble, Association Champollion, pp. 277-292.

WHITELOCK, P. J., WOOD, M. M., CHANDLER, B. J., HOLDEN, N. et H. J. HORSFALL (1986) : «Strategies for Interactive Machine Translation: The Experience and Implications of the UMIST Japanese Project», *Coling '86*, Bonn, pp. 329-334.

ZAJAC, Rémi (1986) : *Études de possibilités d'interaction homme-machine dans un processus de traduction automatique*, Thèse de doctorat, Grenoble.

ZAKI Abu Bakar Ahmad et Hapidah Muhayat NOOR (1991) : «Malay Official Letters Translation», *Proceedings of the International Conference on Current Issues in Computational Linguistics*, Pénang, pp. 413-420.


MODÈLE D'INTENTION
CONNAISSANCES SUR LE DOMAINE
SORTIES
GÉNÉRATION DU TEXTE SOURCE
GÉNÉRATION DU TEXTE CIBLE
EXEMPLES
L1
L2
L3
Ln
MONITEUR DE DIALOGUE
ALGORITHME D'APPAREILLEMENT
système
RÉPONSES
utilisateur
ENTRÉE DE BROUILLON