

# PROBLEMES SCIENTIFIQUES INTERESSANTS EN TRADUCTION DE PAROLE

*Interesting scientific problems in speech translation*

**Christian BOITET**

GETA, CLIPS, IMAG (UJF & CNRS),  
385 rue de la Bibliothèque, BP 53  
38041 Grenoble Cedex 9, France  
Christian.Boitet@imag.fr

## Résumé

La composition séquentielle de composants de reconnaissance de parole, traduction automatique et synthèse vocale fournis par des systèmes commerciaux classiques permet d'ores et déjà de construire des systèmes de traduction de dialogues oraux bilingues utilisables pour des conversations personnelles, moyennant une faible interaction et une grande bonne volonté des utilisateurs. Beaucoup de situations demanderaient un bien plus haut niveau de qualité et d'utilisabilité. Les limitations actuelles viennent de problèmes fondamentaux : absence de traitement du contexte, de lien entre pragmatique et prosodie, et de réelle intégration des composants. Nous proposons quelques solutions, que nous cherchons à mettre en œuvre dans le projet C STAR-II. Enfin, la recherche d'applications dérivées comme l'aide à l'interprétariat sur réseau et l'aide à la compréhension active d'une langue étrangère fait apparaître d'autres problèmes intéressants.

**Mots-clés:** TAO de parole, intégration de composants, traitement du contexte, interaction.

## Abstract

The sequential composition of components of speech recognition, machine translation and speech synthesis makes it already possible to propose spoken dialogue translation systems usable for personal conversations, provided the users accept a weak interaction and show a great deal of good will. Many situations would call for a far higher level of quality and usability. The current limitations come from fundamental problems: absence of context processing, of link between pragmatics and prosody, and of a real integration of components. We propose some solutions, which we try to implement in the C STAR-II project. Last, research towards derived applications such as helps for network-based interpretation and for active understanding of a foreign language leads to other interesting problems.

**Keywords:** Speech Translation, Component Integration, Context Processing, Interaction.

## Introduction

La recherche en Traduction Automatique de Parole (TA-P) a commencé en 1986 avec ATR (Kurematsu & al. 1993, Morimoto & al. 1992, Morimoto & al. 1993) puis NEC au Japon, puis aux USA en 1988, en Allemagne en 1992 avec le projet VerbMobil (Stede & al. 1996, Waibel 1996, Waibel & al. 1995)), et seulement en 1995-96 en France et en Suisse. Plusieurs groupes se sont réunis autour du CLIPS pour constituer un partenaire du consortium international C-STAR II (Consortium for Speech Translation Advanced Research) s'occupant du français, CLIPS++ (CLIPS-Grenoble + LATL-Genève + LAIP-Lausanne + LIRMM-Montpellier)

Le but à court terme est de démontrer la faisabilité de systèmes de traduction (presque totalement) automatique de dialogues multilingues finalisés. Des démonstrations dans le domaine de la préparation de voyages sont prévues fin juillet 1999. Chaque partenaire s'occupe de sa langue : reconnaissance de parole, traduction sous forme textuelle vers au moins une autre langue et/ou vers un format interface "pivot" (IF), ainsi que de la synthèse vocale (à partir du texte ou de l'IF).

Nous avons longtemps travaillé en TA, RP (reconnaissance de parole) et SP (synthèse de parole), mais arrivons tardivement à la TA-P. A l'occasion du travail considérable nécessaire pour rattraper ce retard, nous cherchons à ne pas perdre de vue l'intérêt scientifique de l'entreprise. Or, il est apparu que la composition séquentielle de composants de reconnaissance de parole, traduction automatique et synthèse vocale fournis par des systèmes commerciaux classiques permet d'ores et déjà de construire des systèmes de traduction de dialogues oraux bilingues utilisables pour des conversations personnelles, moyennant une faible interaction et une grande bonne volonté des utilisateurs. Faudrait-il alors renoncer à

toute recherche scientifique en TA-P et ne faire que du développement ?

On retrouve ici une situation analogue à celle qui est apparue en TA quand les premiers systèmes, bien que de très faible qualité linguistique, se sont révélés très utilisables pour la veille (militaire, technique, économique). Mais, ici comme en TA de l'écrit, beaucoup de situations demanderaient un bien plus haut niveau de qualité et d'utilisabilité, à commencer par les dialogues finalisés comme ceux liés à la prise de rendez-vous (VerbMobil) ou à la préparation de voyages (C-STAR II). En spécialisant, on arrive, comme en TA du réviseur, à des systèmes de bien meilleure qualité, mais non extensibles.

Les limitations actuelles viennent de problèmes fondamentaux : absence de traitement du contexte, de lien entre pragmatique et prosodie, et d'une réelle intégration des composants. Nous proposons quelques solutions, que nous cherchons à mettre en œuvre dans le projet C STAR-II. Enfin, la recherche d'applications dérivées comme l'aide à l'interprétariat sur réseau et l'aide à la compréhension active d'une langue étrangère fait apparaître d'autres problèmes intéressants.

## 1. Problèmes fondamentaux

### 1.1. Traitement du contexte

Les systèmes actuels traduisent chaque tour de parole sans tenir compte du contexte global, du stade où en est le dialogue, ni du contexte linguistique.

Le contexte global est formé des informations sur les interlocuteurs et la situation en général. Il est évidemment important de savoir qu'un homme ou une femme parle, pour générer les bonnes formes (I am surprised → je suis surpris/surprise). La position respective importe aussi : un agent doit être poli, un client peut l'être moins. Pour les langues comme le japonais, le rapport social influe fortement sur les choix lexicaux. Quant à la situation en général, il s'agit bien sûr de son thème général (situation finalisée comme la fixation de rendez-vous ou la réservation hôtelière, ou situation plus informelle comme une conversation avec les parents de la famille d'accueil d'un enfant séjournant à l'étranger), et aussi d'informations factuelles (noms de personnes, de lieux...).

Connaître le stade du dialogue pourrait guider la reconnaissance et l'analyse. En ouverture, on s'attend aux salutations et aux présentations, et plus après. En fin de dialogue, on attend des formules de politesse presque figées. Entre les deux, la structure dépend du type de dialogue.

Pour un dialogue finalisé, on a une succession d'épisodes correspondant chacun à la réalisation d'une tâche particulière (réserver une chambre, puis un billet de spectacle, par exemple). Chaque épisode fait intervenir des actes de parole plus ou moins prévisibles (questions pour action ou pour information, confirmations, informations, suggestions, remerciements, conclusion...).

Les dialogues non finalisés n'ont guère été étudiés dans ce cadre, leur structure est sans doute plus souple, mais les mêmes éléments apparaissent. Or les formes linguistiques associées aux divers actes de parole diffèrent parfois de peu (en japonais, on peut avoir un énoncé de type XXX-ka pour une question, et XXX-ga pour une affirmation).

Enfin, le contexte linguistique est la mémoire des "centres" possibles, c'est-à-dire des éléments lexicaux susceptibles d'être élidés ou repris anaphoriquement. Or les élisions et les anaphores sont extrêmement nombreuses dans les transcriptions de dialogues (ATR-ITL 1994, Park & al. 1994, Park & al. 1995). Par exemple :

- (1a) Nous avons deux chambres, une sur cour avec WC et l'autre sur rue avec douche et WC.
- (1b) Pour aller à la gare, ne prenez pas la première rue à droite, mais la seconde.
- (2) D'accord, je prends la seconde.

Il convient de bien calculer l'élision pour traduire correctement dans une langue comme l'allemand (chambre → das Zimmer, neutre, rue → Straße, féminin).

Les "centres" sont aussi utiles en génération, pour faire des choix lexicaux cohérents avec ce qui précède, et en particulier si on génère à partir de l'IF. Par exemple, #hotel en IF peut se réaliser en français comme "hôtel", ou "pension", en japonais comme "hoteru", "ryokan", "minshuku", etc. Et si les choix lexicaux sont corrects, on peut construire des ellipses ou des anaphores elles aussi correctes, de façon à obtenir du style parlé.

Notre solution est assez simple, mais utilise le fait que les unités de traduction de systèmes de TA écrits en Ariane (Boitet & al. 1982) peuvent être beaucoup plus longues qu'une phrase. Pour chaque énoncé à traduire, on construit un "texte" Ariane de la forme suivante, avec des marqueurs appropriés :

- contexte global,
- stade du dialogue (liste de stades et d'actes de parole possibles),
- contexte linguistique,

- texte produit à partir du treillis de mots orthographiques fourni par le RP.

Dans ce treillis, certaines ambiguïtés font "éclater" l'énoncé en énoncés alternatifs, d'autres sont factorisées, selon les souhaits du linguiste.

Exemple (artificiel) :

```
<ctxt_glob> <speaker> client <client> Ma-
dame Durand 70 ans <agent> Monsieur
Biedemeyer 52 ans <firme> NTG <topic>
réservation d'hôtel </ctxt_glob>
<ctxt_dial> <stage> épisode central
<past_sp_acts> question-info info requête
<future_sp_acts> oui non question-info
</ctxt_dial>
<ctxt_ling> pension-hotel_NF réserver_VT
chambre_NF cour_NF réserver_VT pen-
sion-régime_NF prendre_VT rue_NF
</ctxt_ling>
<utterance> <alt> d'accord/_encore je
prends/_rends la seconde
<alt> la cour prend la seconde </utterance>
```

## 1.2. Traitement de la prosodie

Il s'agit d'augmenter la qualité de la prosodie produite, en utilisant la structure syntaxique et les informations pragmatiques disponibles lors de la génération. Pour l'instant, la génération produit un texte et efface toutes ces informations. Le composant de SP réanalyse la chaîne produite, trouve les catégories morphosyntaxiques, calcule éventuellement un parenthésage grossier. Pour améliorer la qualité finale, il faut :

- que la SP accepte, outre le format textuel, un format d'entrée où le texte est annoté par des marques produites au cours de la GS (génération structurale et syntaxique) et contrôlant les paramètres disponibles (vitesse, intensité, pauses, forme mélodique ou micromélodique...),
- qu'il y ait une coopération entre les développeurs de la SP et de la GS. C'est une des raisons de la participation du LAIP (Keller 1997, Keller & al. 1996) à CLIPS++.
- qu'on progresse dans la connaissance précise des liens entre la syntaxe, la sémantique, la pragmatique et la prosodie (un des objectifs de recherche du GEOD).

Cette étude a commencé, mais ne donnera sans doute lieu à implémentation qu'au siècle prochain !

## 1.3. Intégration des composants

Du point de vue informatique se pose un problème aigu d'intégration de composants développés avec des machines, des systèmes d'exploitation et des langages de program-

mation différents. Cela est inévitable, même au sein d'un seul laboratoire, étant donnée l'hétérogénéité des problèmes à traiter et des méthodes employées, qui vont du traitement du signal à l'unification en passant par les méthodes markoviennes, les automates d'états finis, les grammaires hors-contexte étendues, et les systèmes de règles de réécriture avec reconnaissance de patrons complexes. De plus, nous sommes confrontés à la nécessité de pouvoir remplacer un composant par un autre de même type, par exemple remplacer un RP de laboratoire comme ECHO ou JANUS par un RP du commerce comme Dragon Dictate ou Voice Type.

Les solutions que nous étudions, et dont plusieurs sont implémentées ou en cours d'implémentation, sont les suivantes :

- utiliser des structures interfaces plus riche pour le passage d'un composant à l'autre.
- construire toutes les données linguistiques des divers composants, souvent procéduraux, à partir de ressources linguistiques déclaratives et "primaires" (bases de données lexicales et grammaticales).
- organiser le système total mieux que par composition séquentielle.

### 1.3.1. Structures interfaces

Une bonne structure interface entre la RP et la TA est un "treillis pondéré temporellement aligné"<sup>1</sup>, (une entrée, une sortie, pas de boucles, informations symboliques ou structures plus complexes sur les nœuds, poids sur les arcs et les nœuds, temps de début et de fin sur chaque nœud). C'est ce que nous utilisons, même si un RP comme JANUS ne peut produire qu'une "trajectoire" en temps quasi-réel, et met 100 fois plus de temps à produire un treillis de (meilleures) possibilités.

Dans une première étape, nous utilisons le RP ECHO de GEOD, qui produisait un treillis de mots phonétiques. Depuis que nous travaillons à une version française de JANUS et utilisons en attendant un RP du commerce (Voice Type d'IBM, peut-être bientôt aussi Dragon Dictate), les nœuds portent des mots orthographiques.

Si le RP avait un "modèle de langage" hors-contexte augmenté, du type HMM-LR comme ATR l'a fait dans le passé, ou GLR(k) comme dans une version de JANUS, les nœuds pourraient aussi porter de petits arbres donnant la structure des groupes simples reconnus par le modèle, et l'analyse linguistique

<sup>1</sup> On devrait plutôt parler de "treille" comme J. P. Desclés, car deux nœuds peuvent avoir deux prédécesseurs maximaux incomparables.

pourrait partir de là et non d'un texte "plat", profitant du travail déjà fait par le RP.

Comme nous l'avons dit plus haut, une bonne structure interface entre la traduction et la synthèse vocale est un texte annoté. Si la GS dispose d'un dictionnaire contenant les prononciations, on peut même envisager que la SP produise directement une transcription phonétique avec les marques morpho-syntaxiques pour les mots "normaux", laissant à la SP le soin de calculer la prononciation des nombres, sigles et autres mots absents de la base lexicale.

### 1.3.2. Architecture

L'architecture séquentielle (en pipe-line) n'est pas excellente en principe, mais se révèle simple et pratique. C'est elle qui est utilisée dans tous les systèmes connus.

De nombreux chercheurs ont proposé des architectures en "tableau noir" (blackboard) et en "agents". Toutes deux semblent difficiles à mettre en œuvre.

L'idée du tableau noir remonte à la période 1970-75 du premier projet ARPA de reconnaissance de parole. Tous les composants utilisent une même structure de données où ils écrivent leurs résultats et lisent ceux des autres. Le tout est coordonné par un gestionnaire d'agenda.

Il y a plusieurs difficultés : (1) la structure commune n'est en général vraiment adaptée à aucun composant, or certains sont très consommateurs de calcul et ont besoin de structures adaptées ; (2) cette organisation exclue l'usage de plusieurs machines coopérant à travers le réseau et suppose plutôt un ordinateur très puissant, voire parallèle ; (3) la mise au point est très délicate, car on ne peut isoler le fonctionnement d'un composant.

L'idée des agents est plus récente, et vient aussi de l'IA. Chaque composant est autonome et les composants coopèrent à la réalisation d'un but commun (la RP), avec ou sans coordination centrale. Pour cela, ils échangent des messages. La modularité n'est guère meilleure, et la mise au point aussi difficile, car il faut contrôler  $n(n-1)$  flots de messages s'il y a  $n$  composants, au lieu de  $2n$  avec le tableau noir.

Nous prônons plutôt une architecture en "tableau blanc" (Boitet & al. 1994b) :

- les composants sont développés indépendamment et intégrés ensuite
- chaque composant est "encapsulé" par un "gestionnaire" qui le fait apparaître comme un serveur et gère les transformations de formats

- un coordinateur central unique appelle les composants sur des fragments de son "tableau blanc",
- ce "tableau blanc" n'est pas visible par les composants. C'est un ensemble structuré de "tranches" ou "canaux", chacun associé à un composant et contenant une image, éventuellement très partielle, de la structure de travail de ce composant.

Dans notre implémentation actuelle, le tableau blanc se limite aux structures évoquées plus haut (contexte global, stade du dialogue, contexte linguistique, sortie du RP, entrée et sortie de la TA, entrée de la SP), auxquelles on ajoute une mémoire "cache" permettant de ne pas refaire les calculs sur un treillis déjà traité. En effet, certains énoncés reviennent plusieurs fois dans un même dialogue.

## 2. Adaptabilité

Revenons sur l'utilisabilité de systèmes produits par assemblage de composants du commerce. Dans l'expérience présentée par M. Seligman du GETA en coopération avec CompuServe au MT-summit'87 à San Diego, puis au colloque C-STAR'98 à Grenoble, les interlocuteurs utilisent deux PC reliés par modem au réseau téléphonique. Sur chacun, le RP est le produit (anglais ou français) de Dragon. L'énoncé reconnu est affiché à l'écran, et l'utilisateur peut l'accepter, le refuser ou l'éditer. Il y a aussi des cas d'échec pur du RP. Après une moyenne de 3 à 4 interactions, le texte accepté et éventuellement révisé est envoyé au serveur de TA du service "Chat" de CompuServe, qui est une version du système d'Intergraph (F-E ou E-F selon le sens) sur laquelle CompuServe a beaucoup travaillé depuis 1995. Le texte produit est envoyé à l'autre interlocuteur, affiché à l'écran, et prononcé par la SP de Dragon sur son PC.

L'utilisabilité est indubitable, pour une utilisation très occasionnelle. On s'amuse des erreurs et limitations de la machine, et on arrive à communiquer. Mais il faudrait diminuer considérablement l'interaction, et donc augmenter la qualité en conséquence, pour permettre une utilisation fréquente ou professionnelle, car le temps humain perdu est une fraction trop importante du temps total. Pour cela, il faut non seulement améliorer les solutions aux problèmes fondamentaux vus plus haut, mais aller vers une certaine adaptation automatique, dans les deux sens, du système à l'utilisateur et de l'utilisateur au système.

Dans le premier sens, les composants de RP sont souvent munis d'un dispositif d'adaptation au locuteur. Au niveau de la TA, rien n'existe encore. La piste la plus prometteuse se situe dans l'adaptation des dictionnaires. On

pourrait pondérer les mots et les sens des mots et faire évoluer ces poids en fonction des sélections passées. Ou encore, pour des systèmes totalement symboliques, réordonner dynamiquement les listes de mots et de sens de mots.

L'adaptation des grammaires est également possible, dans le cas où l'analyse est fondée sur un modèle pondéré (comme celui du projet de banque d'arbres ATR-Lancaster). Mais il n'est pas sûr que le nombre d'énoncés prononcés par un locuteur donné suffise pour produire une amélioration significative.

Dans le second sens, il y a beaucoup à faire, car les utilisateurs s'adaptent rapidement si on leur en donne les moyens. Il s'agit donc de rendre le système "apprenable" et "autoexplicatif". Par exemple, le système pourrait essayer de diagnostiquer les causes d'une mauvaise reconnaissance et proposer des solutions (parler plus près du micro, articuler, parler moins vite ou plus vite...), et donner des exemples de bonne utilisation.

On pourrait aussi l'équiper de boutons permettant d'indiquer le type d'acte de parole qu'on va utiliser, avec une présélection automatique provenant des prédictions du système de contrôle du dialogue. Confirmer par un simple clic guiderait bien mieux l'analyseur et diminuerait le temps de réponse.

La même idée peut être reprise pour le type et le domaine du dialogue : passage d'une partie de dialogue finalisé à une partie informelle, d'une tâche de réservation à une discussion sur le temps qu'il fait...

Tout cela a l'air très simple, mais rejaillit profondément sur la construction du système global et de ses composants. Le plus gros du travail est d'ordre expérimental, mais on tombe aussi sur des problèmes très difficiles comme la découverte de formalismes permettant la définition dynamique de sous-langages "personnels", tant au niveau lexical que grammatical.

### **3. Applications dérivées et problèmes associés**

La recherche actuelle en TA-P vise à produire des traductions complètes d'énoncés oraux, avec un minimum d'interaction, car il s'agit de dialogues. Ce faisant, on néglige la possibilité de construire des applications dérivées moins ambitieuses, faisant appel aux compétences des utilisateurs, et potentiellement très utiles. Nous pensons principalement à l'aide à l'interprétariat sur réseau et à l'aide à la compréhension active d'une langue étrangère. Comme cela s'est produit en TA de l'écrit, la recherche de telles applications dérivées apparaît d'abord comme purement tech-

nologique et applicative, puis fait apparaître d'autres problèmes intéressants.

#### **3.1. Assistance à l'interprétariat**

Il y a de plus en plus de services d'interprétariat téléphonique, et l'interprétariat sur réseau, avec retour visuel et partage de documents (cartes, textes, vidéos...) va sans doute se développer très vite. On peut chercher à aider le travail d'interprétariat proprement dit, et à améliorer les conditions de son exécution.

La partie la plus fatigante du travail de l'interprète est peut-être le stress lexical. Même si l'on s'est préparé à un domaine particulier, des mots peuvent "échapper". En cabine, l'aide d'un collègue avec qui l'on se relaie est précieuse. Des discussions avec plusieurs interprètes professionnels nous ont convaincus que ce ne serait pas une bonne idée de proposer une aide constante, produite par un système de détection de mots ou termes couplé à un dictionnaire et affichant des propositions sur un écran (comme une sorte de fenêtre "Vocabulaire" dynamique). En effet, la charge cognitive est très lourde, et ce serait une distraction. Par contre, on pourrait tout simplement équiper le poste de travail d'un outil permettant de taper un (début de) terme et d'afficher les équivalents trouvés dans un ou plusieurs dictionnaires actifs, avec un filtrage par le domaine. Rien ici que de très classique.

Par contre, on devrait pouvoir augmenter la productivité des interprètes, très sollicités à certains moments et pas à d'autres, en construisant des systèmes leur permettant de partager leur temps entre plusieurs conversations. On peut utiliser la métaphore de la réception de deux délégations parlant deux langues différentes et accompagnées de quelques interprètes. Beaucoup de conversations se déroulent sans leur aide, les interlocuteurs utilisant leur connaissance de l'autre langue, même imparfaite, ou une autre langue commune. Certaines conversations se trouvent cependant bloquées, et un interprète vient à la rescousse, de lui-même ou sur demande.

La version "virtuelle" de cette situation mène à la conception d'un poste de travail de l'interprète montrant les conversations en cours, sans doute une par écran, affichant leurs paramètres et diagnostiquant diverses difficultés ou l'arrivée d'un blocage, et signalant bien sûr les appels directs. Chaque conversation pourrait bien sûr être aidée par des dispositifs automatiques, allant de dictionnaires à recherche vocale ou écrite à un système complet de TA-P.

Un troisième type d'application dérivée pourrait être un système d'interprétariat multilingue, réalisant sur le réseau le correspondant d'une salle d'interprétariat dans un grand

organisme international. Il s'agirait donc de téléconférence multilingue. Le simulateur que nous construisons actuellement pour constituer une base de données de dialogues multilingues et préparer l'environnement de démonstration du projet C-STAR II est un prototype de ce genre de système, puisque chaque composant de TA-P (pour chaque paire de langues) est remplacé par un "magicien d'Oz" humain qui en joue le rôle. Se posent ici de nombreux problèmes d'architecture de système ainsi que de transmission et de synchronisation de données.

### 3.2. Assistance à la compréhension de langues étrangères

Il s'agit ici de reprendre l'idée qui semble mauvaise pour les interprètes et de l'utiliser pour aider n'importe qui à améliorer sa compréhension orale d'une langue étrangère. Prenons l'exemple de touristes venant en France.

On proposerait un appareil comportant un micro relié à un ordinateur de poche, contenant un reconnaiseur de mots isolés et un dictionnaire bilingue, avec plusieurs niveaux prédéfinis correspondant à des niveaux de compétence en français. À chaque niveau correspondrait le sous-ensemble des mots du français "juste difficiles" pour ce niveau.

On activerait l'appareil en choisissant un domaine particulier (sports, politique, etc.), de façon à obtenir un dictionnaire actif assez petit pour obtenir une détection en temps réel (50—100 mots avec les algorithmes et processeurs actuels). De plus, il serait possible de personnaliser le système en supprimant les mots qu'on juge reconnaître sans peine, et en en ajoutant d'autres. On pourrait ainsi écouter la radio, regarder la télévision, ou mener une conversation face à face, et voir apparaître sur l'écran la ou les traductions des mots détectés, avec éventuellement leur écriture en français.

On voit bien ici que le meilleur axe de travail est dans un premier temps celui de l'accueil et non du voyage, puisqu'on peut utiliser toujours le même reconnaiseur, très lourd à développer, même dans le pays d'origine, et le lier à autant de dictionnaires bilingues qu'on veut, la place mémoire n'étant plus un problème. Ainsi, le même appareil pourrait être proposé à tous les visiteurs étrangers.

## Conclusion

La situation en traduction de parole ressemble actuellement à celle de la traduction de l'écrit à la fin des années soixante-dix. L'assemblage d'outils du commerce permet d'arriver à des systèmes analogues à ceux de TAO du veilleur, utilisables pour des conversations occasionnelles, moyennant le concours actif des interlocuteurs. Pour des utilisations

plus fréquentes, on construit des systèmes finalisés, spécialisés à un type de tâche, analogues aux systèmes de TAO du réviseur. Enfin, on envisage des systèmes d'aide à l'interpréariat individuel, à la téléconférence multilingue, et à la compréhension de langues étrangères.

Tous ces types de TA de parole suscitent des problèmes fondamentaux intéressants, et parfois très ardues. Ils posent aussi des problèmes de communication homme-machine proprement dite, et de construction de systèmes s'adaptant aux utilisateurs et permettant aux utilisateurs de s'adapter à eux, ce qui sera peut-être la clef de l'utilisation de ces techniques par le grand public.

## Références

- (ATR-ITL 1994) Transcriptions of English Oral Dialogues Collected by ATR-ITL using EMMI. TR-IT-0029, ATR-ITL, January 1994, 33 p.
- (Bertenstam & al. 1995) The Waxholm system - a progress report. Proc. ESCA Workshop on Spoken Dialogue Systems, Vigso, Denmark, May 1995.
- (Black & Campbell 1995) Predicting the intonation of discourse segments from examples in dialogue speech. Proc. ESCA Workshop on Spoken Dialogue Systems, Vigso, Denmark, May 1995.
- (Blanchon 1995) An Interactive Disambiguation Module for English Natural Language Utterances. Proc. NLPRS'95, Seoul, 4-7 Dec. 1995, vol. 2/2, pp. 550-555, 6 p. (best paper award for the technical content and the presentation)
- (Blanchon & Loken-Kim 1994) Towards more User-Friendly Natural Language Human-Computer Interaction. *94/109*, pp. 17—24.
- (Boitet 1988) Representation and Computation of Units of Translation for Machine Interpretation of Spoken Texts. *Computers and Artificial Intelligence*, 6, pp. 505—546.
- (Boitet & Blanchon 1994a) Multilingual Dialogue-Based MT for monolingual authors: the LIDIA project and a first mockup. *Machine Translation*, 9/2, pp. 99—132.
- (Boitet & al. 1982) ARIANE-78, an integrated environment for automated translation and human revision. Proc. COLING-82, Prague, July 1982, pp. 19—27.
- (Boitet & Seligman 1994b) The "Whiteboard" Architecture: a way to integrate heterogeneous components of NLP systems. Proc. COLING-94, 5—9 August 1994, pp. 243—246, 7 p.
- (Caelen-Haumont 1996) Prosody and disambiguation. Proc. MIDDIM-96 Seminar, Le Col de Porte, 12—14 August 1996, pp. 283-297.
- (Fafiotte & Boitet 1996) An Analysis of the first EMMI-based Experiments on Interactive Disambiguation in the Context of Automated Interpreting Telecommunications. Proc. MIDDIM-96

- Seminar, Le Col de Porte, 12—14 August 1996, pp. 224—237.
- (Fraser & Thornton 1995) Vocalist : a robust portable spoken language dialogue system for telephone applications. Proc. Eurospeech'95, Madrid, September 1995.
- (Gagnoulet & Sorin 1993) CNET Speech Recognition and Text-to-Speech for Telecommunications Applications. Proc. Joint ESCA-NATO/RSG.10 Tutorial and Research Workshop on Applications of Speech Technology, Lautrach, 16-17 September 1993.
- (Hutchins 1986) Machine Translation : Past, Present, Future. Ellis Horwood, John Wiley & Sons, Chichester, England, 382 p.
- (Keller 1997) Simplification of TTS architecture vs. operational quality. Proc. Eurospeech '97.
- (Keller & Zellner 1996) Output requirements for a High-Quality Speech Synthesis System: The Case of Disambiguation. Proc. MIDDIM-96 Seminar, Le Col de Porte, 12—14 August 1996, pp. 300-308.
- (Kurematsu & al. 1993) Important Issues for Automatic Interpreting Telephone Service. Proc. International Symposium on Spoken Dialogue, 10-12 Nov. 93.
- (Lehrberger & Bourbeau 1988) Machine Translation. Linguistic characteristics of MT systems and general methodology of evaluation. John Benjamins, 240 p.
- (Morimoto & al. 1992) A Spoken Language Translation System: SL-TRANS2. Proc. COLING-92, Nantes, juillet 1992, Boitet, ed., ACL, vol. 3/4, pp. 1048—1052.
- (Morimoto & al. 1993) ATR's Speech Translation System: ASURA. Proc. EuroSpeech'93, Berlin, 21-23/9/83, 4 p.
- (Park & al. 1994) An Experiment for telephone versus multimedia multimodal Interpretation: Methods and Subject's Behavior. Technical Report, ATR-ITL, Dec. 94, 15 p.
- (Park & al. 1995) Transcription of the Collected Dialogue in a Telephone and Multimedia/Multimodal WOZ Experiment. Technical Report, ATR-ITL, Feb. 95, 123 p.
- (Sanderman 1996) Prosodic Phrasing. Production, Perception, Acceptability and Comprehension. Thesis, Technische Universiteit, Eindhoven, The Netherlands.
- (Stede & al. 1996) Disambiguation in VerbMobil. Proc. MIDDIM-96 Seminar, Le Col de Porte, 12—14 August 1996, pp. 276-282.
- (Steeneken & Lamel 1994) SQALE : Speech Recognizer Quality Assessment for Linguistic Engineering. Proc. ARPA Spoken Language Technology Workshop, March 1994.
- (Vauquois & Boitet 1985) Automated translation at Grenoble University. Computational Linguistics, **11**/1, January-March 85, pp. 28—36.
- (Waibel 1996) Translation of Conversational Speech. **29**/7, pp. 41-48.
- (Waibel & al. 1995) Multimodal Interfaces. Artificial Intelligence Review.
- (Wehrli 1992) The IPS System. Proc. COLING-92, Nantes, 23-28 July 1992, Boitet, ed., vol. 3/4, pp. 870-874.
- (Woszczyna & al. 1993) Recent Advances in JANUS : a Speech Translation System. Proc. EuroSpeech'93, Berlin, 21-23/9/83.

-0-0-0-0-0-0-0-0-0-0-