

Interprétariat à distance et collecte de dialogues spontanés bilingues, sur une plate-forme générique multifonctionnelle

Georges Fafiotte

GETA, CLIPS-IMAG (UJF - Université Grenoble 1)
385, rue de la Bibliothèque, BP 53, F-38041 GRENOBLE Cedex 9 (France)
georges.fafiotte@imag.fr

Résumé — Abstract

Parallèlement à l'intégration du français en TA de Parole multilingue (projets C-STAR, NESPOLE!), nous avons développé plusieurs plates-formes, dans le cadre des projets ERIM (Environnement Réseau pour l'Interprétariat Multimodal) et ChinFaDial (collecte de dialogues parlés spontanés français-chinois), pour traiter différents aspects de la communication orale spontanée bilingue non finalisée sur le web : interprétariat humain à distance, collecte de données, intégration d'aides automatiques (serveur de TA de Parole utilisant des composants du marché, interaction multimodale entre interlocuteurs, et prochainement aides en ligne aux intervenants, locuteurs ou interprètes). Les corpus collectés devraient être disponibles sur un site DistribDial au printemps 2004. Ces plates-formes sont en cours d'intégration, en un système générique multifonctionnel unique ERIMM d'aide à la communication multilingue multimodale, dont une variante s'étendra également à la formation à distance (*e-training*) à l'interprétariat.

In parallel with integrating the French language into multilingual Speech Machine Translation (within the C-STAR and NESPOLE! projects), we have developed in recent years several platforms, in the framework of projects ERIM (Network-based Environment for Multimodal Interpreting) and ChinFaDial (collecting French-Chinese spontaneously spoken dialogues), allowing to handle various aspects of spontaneous, general-purpose bilingual spoken dialogues on the web: distant human interpreting, data collection, integration of machine aids including server-based speech translation based on commercial products, multimodal user interaction, and next, online aids to speakers and/or interpreters. Collected data should be available on the web (DistribDial) in spring 2004. All platforms are being integrated into one single multifunctional ERIMM generic system, which should then be extended to distant *e-training* in interpreting.

Mots-clés — Keywords

Interprétariat à distance sur réseau, collecte de corpus oraux bilingues, dialogues spontanés, communication multilingue, mutualisation de ressources.
Web-based interpreting, bilingual spoken corpora collection, spontaneous dialogues, multilingual communication, resource mutualization.

Introduction

Les progrès de la TA de Parole (TAP) ont été sensibles dans la dernière décennie, particulièrement en spontanéité du style et en multilinguisme. La première démonstration fut effectuée par NEC (septembre 1992) dans le domaine du tourisme, mais les recherches les plus connues sur ces thèmes peuvent être créditées à des actions très coordonnées, entre autres les

projets C-STAR (international Consortium for Speech Translation Advanced Research) [8, 10], en Europe les projets NESPOLE! IST [13] et l'allemand Verbmobil [15] et aux USA le DARPA Communicator project [11] avec le Galaxy Communicator Software Infrastructure [12]. Ces projets ont produit des plates-formes de traitement de la parole spontanée en communication multilingue personne-personne ou personne-système, toujours dans des domaines ciblés.

Le manque notoire de grands corpus de dialogues parlés bilingues en accès libre (ressources essentielles au développement de systèmes de TAP spontanée), d'une part, et d'autre part l'importance de l'étude de l'impact de la multimodalité sur la communication multilingue (à l'heure où les équipements de télécommunication mobile monolingue intègrent ce type de dispositif), nous ont motivés pour étudier, modéliser, et prototyper un ensemble de ressources génériques et de plates-formes orientées vers les aides à la communication orale multilingue et multimodale sur le web, en contribution également à certaines attentes de l'ingénierie linguistique.

Cet article rappelle d'abord nos premiers travaux dans cette orientation, puis présente la famille des plates-formes ERIM (Environnement Réseau pour l'Interprétariat Multimodal) en précisant leurs motivations, leurs caractéristiques de conception, et leur état actuel. Il rend compte ensuite des premières utilisations en collecte de dialogues parlés spontanés français-chinois en domaine finalisé (projet ChinFaDial), et aborde les développements en cours et à venir, en vue d'une intégration de ces composants sur une plate-forme unique.

1 Aider la communication multilingue sur réseau

1.1 Situation, enjeux

Les motivations de cette recherche sont multiples : elles sont nées de l'observation des conditions de l'intégration du français aux projets de TAP C-STAR II puis NESPOLE!, et du constat de l'absence de grands corpus de parole spontanée en dialogues bilingues interprétés, essentiels pour la production de modèles linguistiques des langues parlées, très différents de ceux des dialogues monolingues et des textes écrits. Il est, pour chaque langue, de multiples variantes en langue parlée spontanée, tant y sont fréquents les variations de style, de traits syntaxiques particuliers (anaphores, style indirect...), les phénomènes d'élocution (faux départs, interruptions et reprises, raccourcis...), les tournures idiomatiques ou usuelles orales. L'incidence de ces traits de parole spontanée s'avère de plus différente en situation multilingue.

Avec pour finalité une aide à l'ingénierie linguistique, nous souhaitons aussi faciliter une expérimentation fine de composants de TAP (de reconnaissance, traduction, synthèse), et celle de ressources multimodales dans un cadre de traduction de parole bilingue — donc, créer une plate-forme générique, ouverte à l'intégration (plug-in) de composants, facilitant la capture de données.

Nous voulons de plus proposer aux interprètes humains des scénarios innovants et de nouvelles modalités d'intervention sur le web, permettant le travail à distance, par exemple pour l'insertion professionnelle des handicapés, et facilitant les interventions ponctuelles à la demande.

Enfin, nous mettrons à disposition à terme certains des logiciels développés, en accès libre sur le web, pour favoriser un volontariat de production de ressources partageables au sein de la communauté des chercheurs en TA de Parole.

1.2 Un premier environnement : Sim*, simulateur de traduction avec Magicien d'Oz, pour la collecte de dialogues spontanés

Motivation. À la suite de travaux sur des plates-formes multimodales à magicien d'Oz (architecture multimagiciens monolingues de NEIMO [4], ou, à ATR-ITL Kyoto, EMMI [7, 5] à magicien unique bilingue, dans un contexte d'étude de la désambiguïsation interactive multimodale), nous avons pensé utile, pour faciliter la construction de systèmes de TAP, d'acquérir d'abord une expérience et de rassembler des données, en prototypant Sim* (prononcer Sim-star), environnement de simulation de TAP conçu parallèlement à C-STAR II.

Conception. L'architecture de Sim* est celle d'un environnement multiposte avec

interprète magicien d'Oz [6], ce dernier entendant et voyant les locuteurs grâce à leur webcam, et n'étant ni vu, ni entendu (du moins comme un humain) par eux. Sur Sim*, les locuteurs peuvent s'entendre et se voir, et disposent de ressources multimodales de base : échange de textes courts, "tableau blanc" pour le partage de documents visuels et d'annotations graphiques libres. Nous visions une observation de comportements d'utilisateurs placés en situation future de TA multimodale, en même temps que la collecte de données (parole et textes courts dans un premier temps) permettant la modélisation de la langue spontanée parlée (par exemple pour C-STAR II).

Pour ce premier système, la conversation sur réseau est gérée par un serveur de communication de type client-serveur programmé en Tcl/Tk, et l'interaction multimodale utilise des applications Mbone. Implémenté sur stations Unix, Sim* a été utilisé sur l'intranet du laboratoire, bien qu'écrit pour tourner sur Internet. Les tests ont été effectués en anglais-français.

Situation actuelle. La réalisation d'un "magicien d'Oz Interprète" crédible, non reconnu comme humain par les locuteurs, donc utile pour l'observation de ces derniers, s'est avérée difficile même en utilisant des vocodeurs pour déformer la voix de l'interprète. Nous l'avons donc écartée, et avons choisi, pour les situations de collecte comme d'expérimentation multimodale, que l'interprète soit perçu comme tel —un "warm body", ou "ange gardien". Ceci nous permet d'explorer directement des situations réalistes d'usage de systèmes de TAP futurs, tels que nous les concevons : médiatisés et synergiques, c'est à dire intégrant des ressources de traduction "par la machine" (cf. 2.3) et "aidée par l'utilisateur" (cf. 2.1), ce dernier disposant d'aides en ligne (cf. 2.4). Mais la collecte de corpus spontanés restait l'enjeu premier.

Cette conception a été développée lors du prototypage par étapes d'une famille de composants ciblés, qui constituent le système ERIM, et sont présentés dans la section suivante.

2 Les composants d'ERIM, pour l'aide à la communication orale multilingue

La famille des plates-formes ERIM comprend, actuellement : ERIM-Interprète pour l'interprétariat multimodal à distance (et base des autres composants), ERIM-Collecte pour la collecte de corpus de dialogues parlés spontanés bilingues traduits, ERIM-TA (vers un service de Traduction partiellement Automatique de Parole aidée par le locuteur, et banc d'essai de composants de TAP) et ERIM-Aides (vers des aides en ligne à la communication multilingue sur réseau). Elle intégrera ERIM-Formation pour l'e-training à distance en interprétariat bilingue.

2.1 ERIM-Interprète, pour l'interprétariat multimodal à distance

Motivation. Il existe, sur le marché, des environnements "propriétaires" fonctionnant sur réseau, de type cabine d'interprète, analogues aux environnements de traduction fixes destinés aux grandes conférences multilingues (ONU, Communauté Européenne). Mais leur code n'est pas accessible, pour des développements orientés vers la recherche.

De plus, nous envisageons 2 scénarios, différents de ceux de l'interprétariat classique :

- téléconférence ("conference call") : les interlocuteurs prennent un rendez-vous avec un interprète pour un créneau de durée donnée,
- "interprétariat intermittent à la demande" : les locuteurs essaient de converser en utilisant la connaissance qu'ils ont de la langue de leur interlocuteur, ou dans une langue véhiculaire, ou connue des deux. Lorsque cette communication s'avère impraticable, ou pour des séquences "sensibles" de leur échange, ils font appel momentanément aux services d'un interprète disponible sur le web, qui peut les aider.

Parallèlement à la modélisation et au prototypage de ressources orientées vers des services d'interprétariat à distance, une autre motivation pour ces plates-formes est l'étude expérimentale de l'incidence de diverses combinaisons de ressources multimodales, sur les dialogues bilingues ou multilingues. Cette approche expérimentale requiert des fonctionnalités de capture et d'enregistrement de données, à l'origine aussi d'ERIM-Collecte.

Conception. La plate-forme ERIM-Interprète est le socle de l'environnement ERIM, et gère la communication entre interlocuteurs. L'architecture générique, commune avec celle d'ERIM-Collecte, est présentée sur la figure 2 : elle comprend un serveur de communication, deux stations locuteurs, une station interprète. Un serveur de multimodalité la complète pour la communication vidéo et par tableau blanc partagé. Les locuteurs s'adressent le plus souvent à l'interprète, mais peuvent se parler directement s'ils le souhaitent. En multimodalité, sont possibles l'échange de textes courts, le partage sur un tableau blanc de documents textuels et graphiques avec pointages, marquages et soulignements libres, et la vision du correspondant.

Situation actuelle. L'implémentation en Tcl/Tk est multiplate-forme (indépendante des systèmes d'accueil, Windows, MacOS, ultérieurement Linux) et générique. Elle permet des configurations évolutives : le serveur de communication sur station séparée (ou non), 2 stations locuteurs (ou plus) distantes, 1 station interprète distante (ou plusieurs) ; 2 processus interprètes sont possibles sur une même station, par exemple en situation de bilinguisme avec traductions "symétriques" ; 2 processus locuteurs également, en situation de "visite" d'un locuteur à l'autre. Nous adoptons un schéma de type "pousser/parler" (push to talk) pour discipliner les échanges.

Les premiers tests ont été faits à moyenne distance (Grenoble-Valence) :

<u>Exemple d'expérimentation</u> (110 km, à 100 Mbits), avec cotation de 0 à 5	Texte court	Voix RtS : Record-then-Send	Voix S&R : Send & Record (streaming)	Voix S&R + recouvrement de tours de parole
Streaming	non	non	oui	oui
Qualité de réception	5	5	3	1
Rythme des échanges	5	2	4	5
Fiabilité de transmission	5	5	4	1
Phénomènes particuliers		utilisateur hésitant, dialogue perturbé (trop lent)	quelques micro- coupures, bonne qualité générale	inutilisable (alors qu'acceptable sur un intranet)

Fig. 1 : Communication orale sur le web

La mise au point initiale a été faite sur l'intranet du laboratoire. Sur Internet l'échange de messages textuels courts est géré classiquement, mais pour la transmission de la voix deux modes peuvent être choisis (cf. Fig. 1) : soit "enregistrer-puis-envoyer" (Record-then-Send, enregistrement local puis transmission du fichier enregistré), très fiable mais qui crée des temps d'attente parfois contraignants, soit "envoyer-et-enregistrer" (Send & Record, avec transmission en flux —streaming), qui assure un rythme d'échanges satisfaisant et dont la qualité est bonne (malgré quelques microcoupures) pour autant que les locuteurs se parlent sans recouvrement (deux locuteurs ne s'adressant pas en même temps à l'interprète, par exemple).

Des validations à longue distance sont prévues, avec étude des performances en fonction des liaisons utilisées.

2.2 ERIM-Collecte, pour la collecte sur réseau de dialogues spontanés bilingues traduits

Motivation. L'importance de grands corpus réalistes est essentielle pour la construction de systèmes de TA de Parole. Ces systèmes requièrent des corpus acoustiques maintenant largement produits à partir de communications sur le web. Il est besoin également, pour établir des modèles de langues parlées en situations réelles, de corpus parallèles d'énoncés transcrits alignés. Peu de ressources de ce type sont produites, ou librement disponibles.

De même, de grands corpus de dialogues spontanés bilingues sont nécessaires pour développer et valider les systèmes de TAP, mais il en existe très peu (NEC, ATR, ELRA et

quelques autres), et aucun n'est en accès libre. Pourquoi sont-ils "propriétaires" ? Parce qu'ils sont coûteux à collecter, et encore plus à transcrire et annoter : les rendre disponibles gratuitement semble déraisonnable si on les a beaucoup travaillés.

Face à cette situation, nous avons développé ERIM-Collecte (cf. Fig 2), pour

- collecter des données "brutes" (que d'autres équipes traiteront ensuite), les plus multimodales qu'il est possible, et à partir de dialogues réels,
- proposer à des volontaires de produire gratuitement ces données ...en l'échange d'un accès libre aux plates-formes ERIM nécessaires et aux services qu'elles proposent.

Conception. ERIM-Collecte est une extension d'ERIM-Interprète, avec enregistrement systématique des actes et données de l'interaction pour tous les participants (deux locuteurs ou plus, un interprète ou plus). L'enregistrement est fait localement lors de la conversation, sur fichiers son au format PCM 16kHz-16bit mono. En fin de dialogue, les descripteurs et fichiers produits localement sont transmis à un serveur de collecte, où ils sont regroupés et structurés.

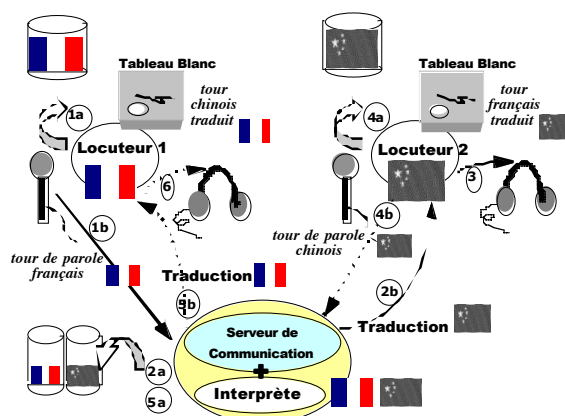


Fig. 2 : ERIM-Collecte, en configuration I locuteurs et i interprètes (ici I=2, i=1)

Dans la situation que présente le schéma, (1) l'interlocuteur français parle (tour de parole en un ou plusieurs énoncés), avec enregistrement local (1a), et transmission au "CommSwitch" qui les diffuse vers un salon virtuel établi pour le dialogue (1b). L'interprète écoute ce tour de parole, le traduit en chinois (2). La traduction est enregistrée localement (2a), en même temps que diffusée (2b). L'interlocuteur chinois écoute la traduction (3), puis répond (4). De nouveau, son tour de parole est enregistré localement (4a) et diffusé (4b). En (5), l'interprète le traduit en français, la traduction étant enregistrée localement (5a), et transmise (5b) au destinataire (6).

Situation actuelle. Sur la version actuelle ERIM/3-Collecte, sont enregistrés les tours de parole et textes courts (bimodalité). La capture des événements du tableau blanc, et des objets impliqués (fichiers visuels partagés, tracés libres, url...) est en cours d'intégration, comme celle de la vidéo, si elle est souhaitée (par exemple pour des études ultérieures d'expressions faciales).

ERIM-Collecte est également réalisée en Tcl/Tk, qui favorise l'utilisation multiplateforme (PC, Macintosh, stations Unix/Linux) sur les stations de travail usuelles des utilisateurs. La figure 3 présente l'interface actuelle de la plate-forme, pour un locuteur.

Une ressource de réécoute (le module Replay) permet de reconstituer tout ou partie du dialogue, chronologiquement ou avec extraction de versions monolingues. Elle facilite un suivi visuel de l'échange (cf. Fig. 4). Elle fonctionne maintenant en version bimodale (tours de parole, échange de textes courts – extension en cours au tableau blanc), et sera accessible sur le site web DistribDial d'accès aux corpus produits, pour des utilisateurs de la communauté scientifique.

Par contre, ERIM-Collecte, destinée par choix initial à la constitution de corpus de données brutes (fichiers descripteurs de session et tours de parole) ne propose pas de ressource particulière d'aide intégrée à la transcription, ni à l'annotation. Elles sont réalisables hors ERIM.

L'architecture ouverte de la plate-forme générique facilitera d'éventuels développements futurs. Nous prévoyons d'utiliser par exemple les reconnaissances vocales intégrées d'ERIM-TA (cf. 2.3) pour produire des premières versions instantanées de transcriptions.

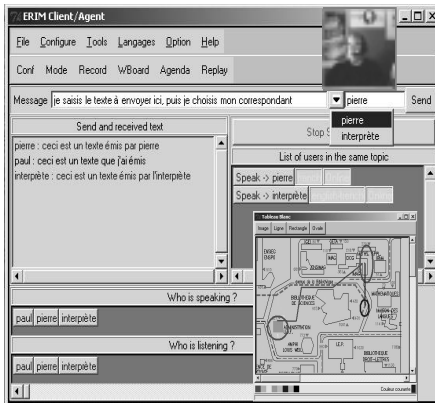


Fig. 3 : Écran Locuteur

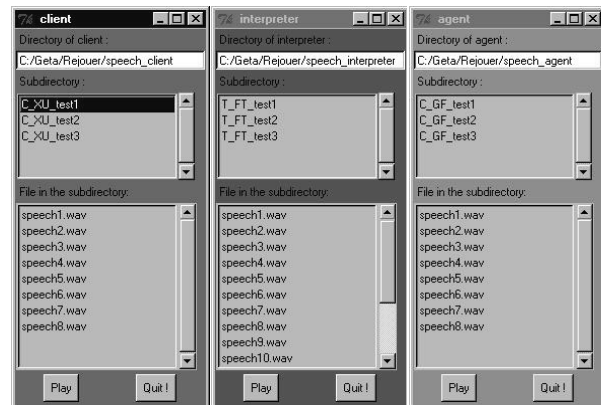


Fig. 4 : Réécoute des énoncés Locuteurs, et Interprète (au centre)

Les versions successives d'ERIM-Collecte ont été utilisées pour les enregistrements, à Grenoble et à Pékin, d'un premier corpus de dialogues spontanés traduits (réservation hôtelière).

2.3 ERIM-TA, vers un service de Traduction partiellement Automatique de Parole (TpAP), aidée par le locuteur

Motivation. Pour cette variante l'objectif était l'intégration générique, par plug-in, de séries de composants de TAP (Reconnaissance, Traduction, Synthèse), pour leur mise au point en évaluation comparative, contrastive, avec la production "humaine" d'un Magicien d'Oz.

S'y est joint, en partenariat avec la start-up Spoken Translation Inc. (STI, Berkeley), le maquetage d'une plate-forme produit que STI souhaite développer. Enfin, nous souhaitons un outil générique permettant d'expérimenter des techniques de Désambiguïsation Interactive dérivées du projet LIDIA [1, 2] et de conduire des expérimentations d'utilisabilité sur ce thème.

L'enjeu global est pour nous une "Traduction partiellement Automatique de qualité", de parole ("tchat") et de texte (SMS), sans recours à un interprète ou traducteur humain, mais en introduisant un niveau adéquat de contrôle par l'utilisateur (systèmes synergiques de TAP).

De tels systèmes devront être utilisables sur tous supports (PC, PDA, téléphones ou futures microstations mobiles). Leurs Reconnaissance Vocale (RV), Traduction Automatique (TA), Synthèse Vocale (SV) et Désambiguïsation Interactive (DI) doivent de plus, au moins pour les équipements mobiles, tourner sur un serveur. Là encore, une généricité des services intégrés est essentielle. L'architecture ouverte et l'intégration différentielle par plug-in y contribueront.

Conception. Pour que les couvertures lexicale et grammaticale soient assez larges, nous avons choisi d'intégrer des composants commerciaux (de reconnaissance de parole, de traduction, et de synthèse de parole) tous disponibles sur serveur. Ce sont, dans le premier prototype, des produits disponibles respectivement chez Philips, Linguatex, et ScanSoft, (avec SDKs, environnements ou kits de développement). Le composant de TA doit pouvoir être couplé ultérieurement avec un module de DI.

Pour la reconnaissance, nous souhaitons une RV de type "dictée vocale", interactive. IBM proposait un SDK pour son composant de RV, offrant une reconnaissance interactive pour des domaines finalisés, et un système de dictée ("transcription") non interactive à large couverture, mais pas de dictée vocale interactive. L'offre de Philips semble à ce jour être la seule permettant une dictée vocale interactive sur serveur, mais l'évolution du domaine est rapide.

Situation actuelle. L'étude d'un banc d'essai monostation —le "rack" d'intégration— a été conduite, avec plug-in des composants externes. Ce prototype a été réalisé en Tcl/Tk (pour ses facilités d'intégration de ressources externes et d'interopérabilité), puis testé. L'intégration finale à un dispositif multistation, de type ERIM-Interprète, est en cours.

Actuellement, le locuteur émet un énoncé de parole en langue source, et l'énoncé textuel "reconnu" lui est soumis pour validation ; si nécessaire, il peut réentrer son énoncé vocalement

(ou s'il préfère, modifier au clavier l'énoncé textuel proposé). Après validation, le composant de traduction produit un texte en langue cible, synthétisé par le composant de synthèse vocale.

Dans certains cas (locuteur émetteur partiellement bilingue) cette traduction écrite peut lui être renvoyée ; une rétrotraduction peut aussi lui être soumise, pour validation ; elle pourra jouer un rôle d'indicateur, certes assez imparfait, d'une non-pertinence de la traduction. En cas de réelle difficulté, le locuteur peut reformuler oralement (voire au clavier) son énoncé en langue-source, pour contourner une possible inaptitude de la TA. Une validation provoquera la synthèse vocale.

Il est prévu qu'ERIM-TA effectue l'enregistrement de toutes les productions (de reconnaissance, de traduction et rétrotraduction, et de synthèse), et de toute reformulation. Les descripteurs correspondants et la base de fichiers d'une session sont en cours d'implémentation.

Une modélisation de la désambiguïsation interactive, fondée sur les techniques LIDIA [1, 2], est prévue en partenariat avec Spoken T. Inc. ; elle sera prototypée, puis intégrée et expérimentée sur la plate-forme ERIM-TA. Nous prévoyons d'effectuer une pré-expérimentation sur les modalités de ces interactions et leur pertinence, dans différentes situations d'utilisation, en collaboration avec MultiCom (qui est une composante du CLIPS, équipe-ressource d'étude de l'utilisabilité et de l'ergonomie des logiciels), avec observation des comportements d'utilisateurs .

Le paradigme d'une TA de Parole "aidée par l'utilisateur" semble a priori se prêter à des utilisations de "tchat" oral multilingue de qualité, qui intéresse d'abord notre partenaire, et plus globalement de communication multilingue avec traduction de qualité, notre objectif final.

2.4 ERIM-Aides, vers des aides en ligne à la communication multilingue sur réseau

Motivation. Dans notre scénario d'interprétariat intermittent "à la demande", il est proposé aux interprètes de passer d'une conversation à une autre et d'un sujet à un autre (à la manière des interprètes "de cocktail"). C'est une activité difficile, et il peut en résulter un "stress lexical". Des aides automatiques en ligne, brèves —par exemple pour une "mini-immersion lexicale" avant intervention, selon la spécificité du domaine concerné—, seraient bienvenues.

On devrait également disposer d'aides automatiques pour les locuteurs (par exemple "faiblement" ou partiellement bilingues, ce qui concrètement est parfois le cas), pour les aider à "se débrouiller sans interprète", en cas de nécessité.

Conception. Les "aides à la communication" comprennent, entre autres, des ressources

- pour se voir en se parlant (locuteurs, interprète),
- pour partager des données, éventuellement modifiables, pointables, marquables (tableau blanc, visuels partagés...), consultables a posteriori...
- pour planifier et gérer des rendez-vous sur un agenda (accessible sur un site serveur).

Des "aides linguistiques" peuvent inclure :

- l'accès à des fiches terminologiques thématiques bilingues, à des dictionnaires électroniques à saisie au clavier ou à entrée vocale, par exemple par détection de mots automatique (word-spotting) suivi d'un filtrage, d'une recherche en dictionnaire, et présentation de synthèse sur fenêtre unique,
- en l'absence d'interprète, une reconnaissance de parole atténuant les difficultés de compréhension orale et produisant une trace ou un historique de la conversation, que peut également consulter l'interprète avant intervention,
- de la TA de Parole partiellement (ou complètement) automatique.

Situation actuelle. Les aides de communication sont prototypées. L'agenda est global sur un site d'accès ERIM, chaque utilisateur y accédant via une vue personnalisée.

Les premières aides linguistiques vont être introduites, en interfaçant des ressources dictionnaires existantes, en accès libre sur le site Papillon [14]. Une ressource de reconnaissance vocale (Philips) a déjà été interfacée avec ERIM-TA, et sera intégrée.

2.5 ERIM-Formation, pour l'e-training à distance en interprétariat bilingue

En projet actuellement (même si l'architecture générique des plates-formes actuelles permet de simuler déjà son fonctionnement), cette variante permettra de proposer, à des étudiants en interprétariat bilingue, différents modes de formation à distance (FAD) sur le web de type e-training, pour une activité "live" (en direct), ou "de doublage".

Nous prévoyons un dispositif de type "laboratoire de langues sur le web, pour interprètes", avec par exemple un interprète professionnel (et/ou professeur) assurant la continuité et la fluidité du dialogue bilingue entre deux locuteurs monolingues sur le réseau, et un cercle d'interprètes "juniors" qui peuvent (sans entendre les traductions du professionnel) s'exercer à distance, ou éventuellement intervenir chacun à son tour sur proposition d'un médiateur de traduction. Toute intervention étant enregistrée (avec l'accord des intervenants) et consultable, un travail pédagogique intéressant peut s'ensuivre.

Cette ressource rejoint de fait l'approche collaborative que privilégie le projet ERIM, car il est probable que des étudiants avancés en interprétariat ou des interprètes en perfectionnement seront volontaires pour coopérer à l'activité collaborative sur ce type d'outil "d'apprentissage à distance par la pratique", contribuant à la création de corpus de dialogues spontanés (cf. 3.2).

Il est possible également d'envisager des situations d'utilisation plus institutionnelles, où des étudiants "seniors" en interprétariat accepteraient d'assurer des traductions bilingues dans le cadre d'un service multilingue d'interprétariat "grand public", bénévolement ou en échange d'une validation académique de cette activité (par exemple lors de Jeux Olympiques, Beijing 2008).

3 Application à la collecte de corpus de dialogues oraux bilingues

3.1 ChinFaDial, un premier corpus de dialogues spontanés traduits

ERIM-Collecte a été utilisé dans le cadre du projet LIAMA "ChinFaDial", en partenariat avec le NLPR (Institut d'Automatique de l'Académie des Sciences de Chine, à Pékin), pour la collecte de dialogues parlés bilingues français-chinois spontanés et traduits, entre deux locuteurs monolingues (cf. Fig. 5). La transcription ultérieure est ici manuelle.

顾客/Client (7)

我在火车站: 火车站离你们旅馆不知道远不远、怎么走?

(Je suis à la gare, je ne sais pas comment me rendre à l'hôtel à partir de la gare.)

Agent/代理 (7)

Alors c'est extrêmement simple, en sortant de la gare vous tournez à droite et c'est à 80 mètres en face de l'autre côté de la place.

(很简单、如果你出了火车站以后向右转、只要走到80米左右。就是我们的旅馆。)

顾客/Client (8)

好谢谢那就一会儿见

(Merci bien, alors à tout à l'heure)

Agent/代理 (8)

Merci, bonsoir Monsieur, à tout à l'heure.

(谢谢...谢谢、那么、这是...一会见)

Fig. 5 : Dialogue français-chinois (extrait), entre hôtelier français et client chinois

Ces collectes utilisent pour le moment, localement, les intranets des laboratoires partenaires, avec trois participants dans le même bâtiment (ou dans deux bâtiments proches). La collecte à longue distance est en cours de validation.

Les situations sont réelles (réservation hôtelière), en dialogue spontané. Une douzaine d'heures de dialogues oraux ont été enregistrées lors de quelques sessions à Grenoble et à Pékin, et constituent un premier corpus de données brutes, non transcrites. Les collectes se poursuivent.

L'AUF (Agence Universitaire de la Francophonie) finance actuellement le projet VTH-Fra.Dial, une action de collecte de dialogues parlés spontanés bilingues entre le français d'une part, et le vietnamien, le tamoul, et l'hindi, dans des régions francophones des pays concernés.

3.2 Un schéma de création collaborative de ressources pour la TAP

Nous souhaitons contribuer à promouvoir un schéma de mutualisation de ressources pour l'ingénierie de la TA de Parole : corpus parallèles de dialogues parlés spontanés, à collecter, puis à transcrire et annoter, en participation collaborative.

Une étape première sera la mise en accès libre au printemps 2004, sur site web, des données déjà collectées en français-chinois, sous une forme "rejouable" avec le module Replay, et, pour chaque conversation, un descripteur des paramètres d'intervenants (anonymes) et de session, la liste des tours de parole horodatés par locuteur, les pointeurs sur les fichiers son.

Ce site et son environnement DistribDial devront faciliter l'enrichissement libre des corpus par d'autres chercheurs, par ajout de transcriptions et/ou d'annotations en fichiers parallèles (selon un format homogène) — à rendre accessibles également sur le web.

Une contribution suivante sera de proposer la plate-forme ERIM-Collecte elle-même, en accès libre (GPL) sur le web, après de nouveaux tests de robustesse, de performance à distance, et quelques développements d'utilisabilité.

4 Développements en cours, prospective

L'environnement ERIM s'est construit par prototypage exploratoire puis développement incrémentiel de plusieurs classes de ressources complémentaires, en cohérence fonctionnelle. La recherche sur et avec ces plates-formes, après cadrage en tranches fonctionnelles, se poursuit par

- de nouvelles collectes dans de nouvelles langues,
- l'enrichissement fonctionnel du serveur de multimodalité d'ERIM-Interprète, puis le développement de la capture de données multimodales par ERIM-Collecte, enregistrant les événements multimodaux et leur horodatage, avec un Replay Multimodal associé,
- la consolidation de DistribDial, le site d'accès aux corpus collectés,
- un site d'accès à la plate-forme ERIM-Collecte pour la collecte collaborative,
- quelques extensions fonctionnelles conduisant à une plate-forme "laboratoire", facilitant le plug-in de composants pour la TAP ou la "TAP de qualité aidée par l'utilisateur", et leur expérimentation — que ces composants proviennent de réalisations académiques, de versions à tester ou à régler ("tuning"), ou qu'ils soient des produits logiciels du marché ; plusieurs produits ou versions de modules de TA en cours de prototypage, d'une même classe fonctionnelle (RV, TA, SV, ou DI) pourront être testés et comparés en parallèle avec enregistrement complet ; des études contrastives avec des productions humaines d'interprètes magiciens d'Oz, enregistrées en parallèle, seront possibles,
- le prototypage et l'évaluation expérimentale de solutions de désambiguïsation interactive, pour ERIM-TA.

Nous souhaitons, parallèlement à la diffusion d'une plate-forme à usage ciblé (pour la collecte de données brutes, puis leurs enrichissements contributifs), entreprendre également l'unification et l'intégration des différents composants d'ERIM ici présentés en une plate-forme unique ERIMM, Environnement Réseau pour l'Interprétariat Multimodal Multilingue, regroupant un ensemble d'aides à la communication multilingue et multimodale sur réseau.

Conclusion

Nous avons présenté 3 plates-formes déjà opérationnelles (l'une d'elle en finition), permettant d'aider la communication bilingue sur le web, pour des dialogues spontanés non finalisés : ERIM-Interprète, pour l'interprétariat humain sur réseau, ERIM-Collecte, qui réduit le

manque de données utiles pour développer de meilleurs systèmes de TA de Parole, et ERIM-TA ressource générique destinée à la Traduction partiellement Automatique de Parole (TpAP), de qualité, utilisant des produits logiciels de reconnaissance, traduction et synthèse à couverture large et disponibles sur serveurs, avec contrôle par l'utilisateur (feedback, validation directe, désambiguïsation interactive). ERIM-TA assiste aussi l'ingénierie des systèmes de TAP, et constitue un banc d'essai générique en situation réelle, ou de réglage, pour leurs composants.

Une autre plate-forme ERIM-Aides proposant aux interprètes et locuteurs des aides en ligne (aides de communication, aides lexicales), est partiellement réalisée.

Les dialogues spontanés bilingues déjà collectés (français-chinois) seront prochainement disponibles en accès libre sur le web, ainsi que les plates-formes Interprète et de Collecte.

Cette recherche se poursuit par la collecte et la distribution de données concernant d'autres langues, par l'enrichissement fonctionnel des quatre premières plates-formes, puis leur unification en un environnement unique ERIMM, base d'expérimentation et d'aide à l'ingénierie de systèmes de TpAP multilingue et multimodale, et enfin par le développement d'un "labo de langue sur le web pour l'interprétariat", ERIMM-Formation, qui pourrait également contribuer aux collectes.

Remerciements

Ces travaux ont été soutenus par le CLIPS-IMAG (Université Joseph Fourier Grenoble-1, INPG, CNRS), par la Région Rhône-Alpes (projet ERIM), et par le laboratoire franco-chinois LIAMA (projet ChinFaDial). Zhai JianShe (Université de Nankin) en résidence au CLIPS, puis Julien Lamboley (Élève-Ingénieur INSA Lyon) ont contribué au développement des plates-formes. L'auteur les remercie, ainsi que les membres du CLIPS, de MultiCom (à Grenoble) et du NLPR (CAS-IA, à Pékin) qui ont participé aux collectes et expérimentations.

Références

- [1] Blanchon H. (1994) Perspectives of DBMT for monolingual authors on the basis of LIDIA-1, an implemented mockup. *Proc. 15th International Conference on Computational Linguistics, COLING-94*, Y. Wilks ed., vol. 1/2, pp. 115—119.
- [2] Boitet C., Blanchon H. (1994) Multilingual Dialogue-Based MT for Monolingual Authors: the LIDIA Project and a First Mockup. *Machine Translation 9/2 1994*, pp. 99—132.
- [3] Brown R. D., Nirenburg S. (1990) Human-Computer Interaction for Semantic Disambiguation. *Proc. COLING-90, ACL*, H. Karlgren ed., vol. 3/3, pp. 42-47.
- [4] Coutaz J., Salber D., Carraux E., Portolan N. (1996) NEIMO, a Multiwork station Usability Lab for Observing and Analyzing Multimodal Interaction. *Proc. CHI'96 companion*.
- [5] Fafiotte G., Boitet C. (1994) Report on first EMMI Experiments for the MIDDIM project in the context of Interpreting Telecommunications. *MIDDIM report TR-IT-0074 GETA-IMAG & ATR-ITL*, Aug. 1994, 11 p.
- [6] Fafiotte G., Zhai J.-S. (1999) A Network-based Simulator for Speech Translation. *Proc. NPLRS'99, Beijing, 5-7/11/99*, B. Yuan, T. Huang & X. Tang ed., pp. 511-514.
- [7] Loken-Kim K.-H., Yato F., Morimoto T. (1994) A Simulation Environment for Multimodal Interpreting Telecommunications. *Proc. IPSJ-AV workshop, March 1994*, 5 p.
- [8] Morimoto T., Takezawa T., Yato F., Sagayama S., Tashiro T., Nagata M. & al. (1993) ATR's Speech Translation System: ASURA. *Proc. EuroSpeech'93, Berlin, 21-23/9/83*, 4 p.
- [9] Nyberg E. H., Mitamura T. (1992) The KANT system: Fast, Accurate, High-Quality Translation in Practical Domains. *Proc. COLING-92, ACL*, vol. 3/4, pp. 1069—1073.
- [10] <url> site web C-STAR: <http://www.c-star.org>
- [11] <url> site web DARPA: <http://www.darpa.mil/ito/research/com/index.html>
<http://fofoca.mitre.org/doc.html>
- [12] <url> GALAXY architecture site: <http://www.sls.lcs.mit.edu/sls/whatwedo/architecture.html>
- [13] <url> site web NESPOLE! : <http://nespole.itc.it>
- [14] <url> site web PAPILLON: <http://www.papillon-dictionary.org>
- [15] <url> site web VERBMOBIL: <http://verbmobil.dfki.de>