

Outils pour la création de ressources linguistiques et le développement d'aides à la communication multilingue sur le web

Par Georges FAFIOTTE
Laboratoire CLIPS-IMAG
Université de Grenoble I (France)
georges.fafiotte@imag.fr

Par Christian BOITET
Laboratoire CLIPS-IMAG
Université de Grenoble I (France)
christian.boitet@imag.fr

Le développement récent d'outils de téléphonie et de visioconférence ouverts aux transactions sur le réseau Internet suscite l'essor d'activités nouvelles de commerce électronique, de téléservice, avec dialogue libre ou de négociation, demande de renseignements, recherche d'informations, etc. Maintenant à l'oral, comme à l'écrit, le multilinguisme devient un enjeu clef de ces interactions à distance, avec comme questions sensibles un équilibre à construire entre les grandes langues véhiculaires, parmi lesquelles le français, et également un réel maintien de la diversité des langues d'origine des conversants.

Dans notre recherche en conception et production d'outils et de ressources d'aide à la communication multilingue sur le web, et parallèlement à notre effort d'intégration du français en Traduction Automatique (TA) de Parole multilingue (projet international C-STAR en allemand-anglais-coréen-chinois-français-italien-japonais, ou projet européen quadrilingue NESPOLE!), nous avons développé plusieurs plates-formes génériques en choisissant une approche différente de celle de la TA, dans le cadre des projets ERIM (Environnement Réseau pour l'Interprétariat Multimodal) et ChinFaDial (collecte de dialogues parlés spontanés français-chinois).

Ces systèmes permettent de traiter différents aspects de la communication orale spontanée bilingue non finalisée, sur le web : ce sont, chronologiquement, l'interprétariat humain à distance, la collecte de corpus, l'interaction multimodale entre locuteurs, l'intégration partielle d'aides automatiques (serveur de TA de Parole utilisant des logiciels du marché), et prochainement des aides en ligne aux intervenants, locuteurs ou interprètes.

De fait, l'arrivée de la multimodalité dans les équipements de télécommunication mobile (c'est à dire en communication parlée, mais avec échange d'informations écrites brèves, d'images, de schémas, avec tracés libres sur documents graphiques partagés, marquage dynamique des informations visuelles, désignation tactile...) appelle aussi, en recherche, la création de plates-formes d'expérimentation sur de telles ressources pour une étude systématique des facteurs ergonomiques et cognitifs, en d'autres termes pour évaluer l'impact du "multimodal" sur la traduction assistée de conversations spontanées bilingues, dans différents contextes d'application.

Les corpus de dialogues parlés spontanés, bilingues et traduits, qui ont été à ce jour collectés (dans le contexte du projet ChinFaDial, une douzaine d'heures en français-chinois), devraient être disponibles sur un site DistribDial du GETA-CLIPS au printemps 2004, pour des recherches académiques en linguistique, pour alimenter notre ingénierie de la Traduction Automatique de Parole, ou ultérieurement en didactique des langues. D'autres corpus doivent être collectés en français-viennamien, français-tamoul (à Pondichéry), français-hindi, dans le cadre du projet VTH-Fra.Dial soutenu par l'AUF (réseau LTT). Des actions complémentaires pourraient concerner la collecte contrastive de parlers régionaux, avec éventuelle médiation d'un interprète.

Les deux plates-formes ERIM-Interprète —d'aide à l'interprétariat humain à distance sur le web—, et ERIM-Collecte —pour la collecte de corpus de dialogues parlés bilingues traduits—, seront également mises à disposition en accès libre sur un site web, après des tests complémentaires de robustesse et d'utilisabilité, pour une utilisation collaborative, ou pour toute activité de recherche ou de préservation de patrimoine linguistique.

ERIM-TA assiste l'ingénierie de la TA de Parole, en ajoutant l'activation différentielle de composants de reconnaissance vocale, de traduction, et de synthèse de parole, issus de la recherche ou de logiciels commerciaux.

Une plate-forme ERIM-Aides (en cours) permet l'investigation d'aides en ligne issues de la technologie de la TA, et destinées aux interprètes humains intermittents travaillant, à la demande, à distance sur le web : aides disponibles, avant leur intervention (consultation de ressources terminologiques bilingues ciblées), pendant celle-ci, malgré la forte charge perceptive et cognitive de leur activité (par exemple, dans leurs intervalles d'attente, consultation de traces écrites d'énoncés produites par reconnaissance vocale des langues utilisées, avec pointage de mots à la volée, pour une consultation instantanée de mémos terminologiques bilingues, pour la capture au vol de termes sensibles sur un bloc-notes de session), ou encore après la session de traduction (création d'un bloc-notes thématique personnel, par capture d'énoncés oraux ou textuels pointés dans les historiques de sessions).

Une variante ERIM-Formation s'étendra prochainement à la formation à distance à l'interprétariat (un "e-training"), avatar d'un "laboratoire de langues sur le web", pour interprètes en formation ou en perfectionnement, permettant également l'investigation de nouvelles activités d'Apprentissage des Langues à Distance.

Ces différentes plates-formes suivent par ailleurs un processus d'intégration sur une plate-forme unique ERIMM, d'aide à la communication "Multilingue et Multimodale" sur réseau.

Il nous apparaît que de tels outils (corpus collectés, comme plates-formes d'interprétariat et de collecte), et leur mise à disposition prochaine sur des sites en téléaccès libre, au delà des services de communication multilingue sur le web, pourraient sous-tendre des efforts de développement de ressources linguistiques, tant pour des recherches fondamentales ou des actions de préservation de patrimoines linguistiques, que pour développer l'apprentissage de l'interprétariat bilingue, ou pour des actions facilitant l'apprentissage des langues à distance.

Fig. 1 : ERIM-Interprète/Collecte
(collecte de dialogues oraux spontanés français-chinois traduits)

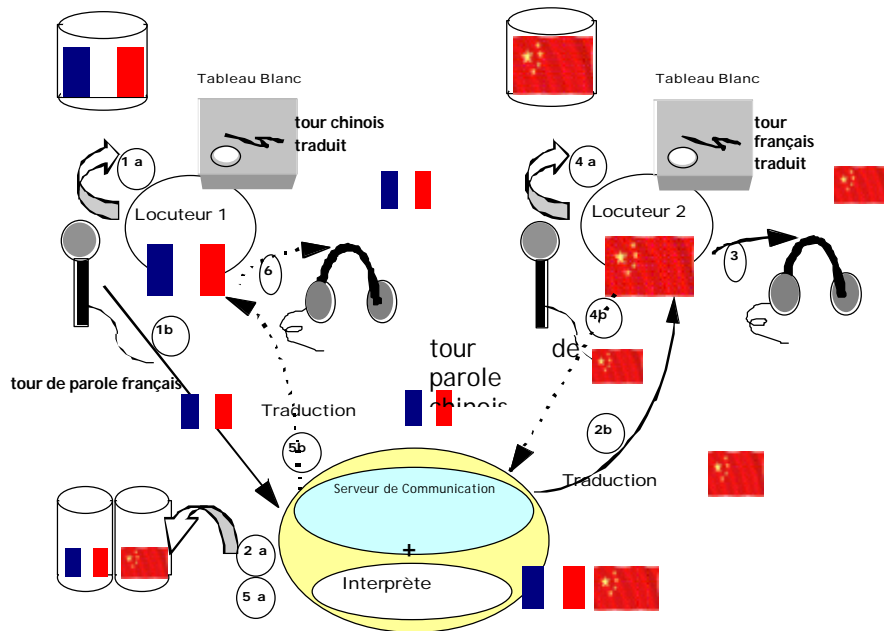


Fig. 2 : Poste de travail d'un locuteur, sur la plate-forme ERIM-Interprète/Collecte (version ERIM/2)

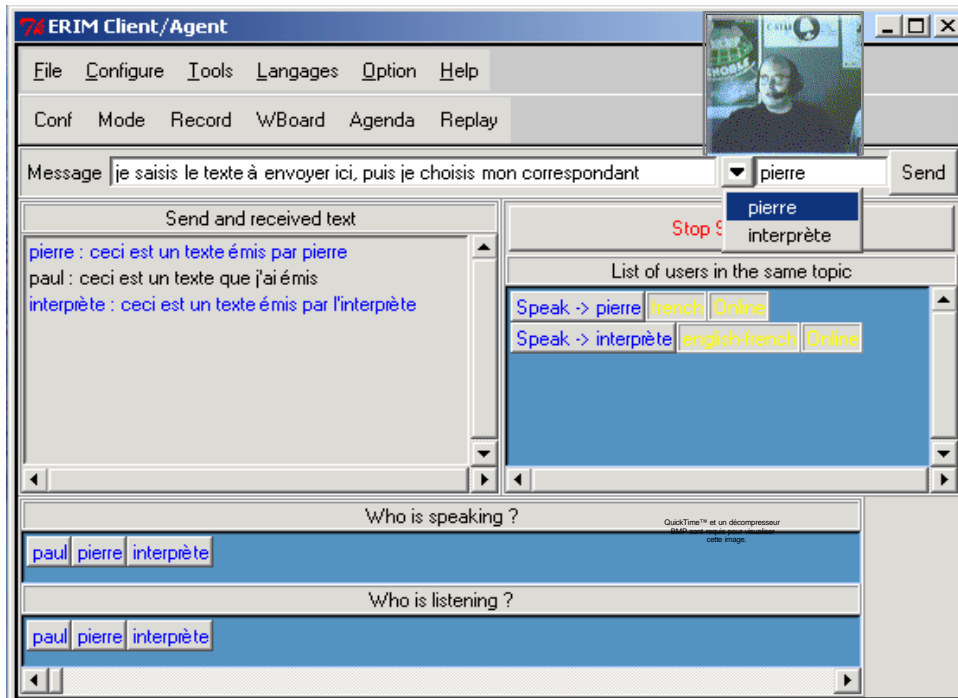


Fig. 3 : Nouveaux scénarios, pour l'interprétariat intermittent à distance sur le web

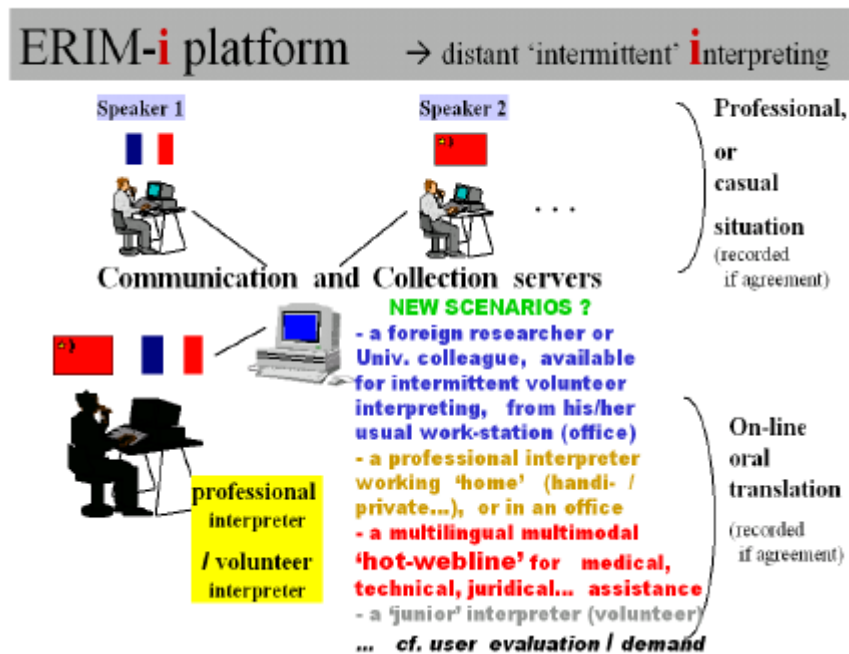


Fig. 4 : ERIM-Formation, pour l'e-Training en interprétariat bilingue, à distance, sur le web

