



Reconnaissance automatisée de la parole : de la parole aux actes... et aux sous-titres

La reconnaissance automatisée de la parole contribue désormais à l'efficacité du sous-titrage et de la transcription, au profit des sourds et malentendants, grâce à la technologie développée par le CRIM. La communauté scientifique canadienne a reconnu la valeur de ces recherches en attribuant à Pierre Dumouchel, vice-président scientifique de cette organisation, le prix IWAY 2004, remis par CANARIE, dans la catégorie des technologies destinées aux personnes handicapées.



Par Jean-François Ferland

Les communications publiques ne sont pas faciles pour un sourd ou un malentendant, qui doit souvent recourir à un interprète gestuel ou attendre la transcription pour bien comprendre une situation. À la télévision, le nombre d'émissions sous-titrées augmente sans cesse, mais les sous-titres sont encore trop rares dans les émissions en direct. Toutefois, l'évolution de la reconnaissance de la parole permet d'automatiser ces tâches de façon fiable, ce qui donne de l'espoir aux personnes en quête d'information.

Les récents travaux de R-D en informatique et en analyse sémantique du CRIM ont conféré une maturité accrue à la reconnaissance automatique de la parole. Les recherches des trois dernières années en adaptation des modèles acoustiques, linguistiques et statistiques, en création d'algorithmes et en développement de dictionnaires de mots ont augmenté le niveau de fiabilité des applications. Ces avancées technologiques ont été mis à contribution dans deux projets novateurs qui permettent aux malentendants, tout comme aux « bien entendants », d'avoir accès à des contenus oraux transcrits en temps réel.

La parole est au député...

Le premier projet visait, comme son nom l'indique, la reconnaissance automatique de la parole, la transcription automatique et l'accessibilité universelle aux débats du Parlement et aux témoignages présentés aux comités (RAP). Cette application destinée aux travaux des **Communes**, du **Sénat** et des commissions avait pour but de faciliter la communication aux sourds

et aux malentendants et d'optimiser la reproduction des contenus en format texte, à des fins de production et de diffusion.

Piloté par le CRIM et soutenu par **CANARIE**, le **Conseil national de recherche du Canada**, **Inno-centre** et des partenaires privés québécois, le projet a mené à la réalisation d'une application multitâche qui facilite le sous-titrage et la transcription de la parole en temps réel, sur place ou à distance. L'application repose sur un serveur de reconnaissance vocale qui analyse les mots prononcés afin de les reproduire en caractères d'imprimerie. Elle a notamment recours au langage XML, à des corpus de données évolutifs et à des objets d'apprentissage associés qui améliorent sa précision au fil du temps.



Pierre Dumouchel,
Vice-président scientifique
CRIM

Pour le sous-titrage, en moins d'une seconde, les mots d'un locuteur qui répète tous les propos dits lors d'un débat sont automatiquement analy-

sés, transcrits, diffusés et affichés sous forme de sous-titres sur les téléviseurs. Les sous-titres peuvent être transmis à un assistant numérique sans fil qui permet à une personne sur place de lire les transcriptions dans le creux de sa main.



Richard McNicoll,
Directeur général
Regroupement québécois
pour le sous-titrage (RGST)

La production des hansards (procès-verbaux), des journaux officiels et des témoignages est pour sa part accélérée par la transcription semi-automatique des locutions. Le transcrip-teur peut ainsi se concentrer sur la révision plutôt que sur la transcription intégrale. Pour les archives du Parlement, un utilitaire d'indexation assistée produit des résumés et extrait des mots clés des documents transcrits. Ces derniers sont ensuite validés par des indexeurs, alors qu'un moteur de recherche évolué fournit des résultats plus précis.

Pierre Dumouchel, vice-président scientifique du CRIM et professeur à l'École de technologie supérieure, affirme que le degré de fiabilité de la reconnaissance de la parole se doit d'être très élevé, afin d'offrir un gain réel en efficacité aux organisations qui y ont recours. Une tâche automatisée doit nécessiter le moins de corrections possible, pour que l'économie de temps soit significative.

« Nos expérimentations sommaires nous démontrent qu'un taux de précision de 70 % lors de la transcription semi-

automatisée nécessite le même temps que la traduction manuelle. Par contre, un taux de 95 % représente une économie de temps de 40 % »

Et comment se compare la performance de la machine avec celle d'un humain ? « Selon une étude de l'université américaine Purdue, un transcrip-teur qui n'utilise aucun système de reconnaissance et qui a tout son temps obtient un taux variant entre 81 % et 95 %, explique le chercheur. Présentement, les performances du système de reconnaissance du CRIM sont de l'ordre de 80 %, et nous visons à atteindre 95 % d'ici deux ou trois années. »

L'application RAP, dont la démonstration devant le Parlement en mars 2004 a été fort appréciée par les sénateurs et les députés, suscite l'intérêt d'autres gouvernements provinciaux. De plus, elle a valu à Pierre Dumouchel, en novembre 2004, le prestigieux prix IWAY décerné par CANARIE dans la catégorie des technologies destinées aux personnes handicapées.

Selon **Richard McNicoll**, directeur général du **Regroupement québécois pour le sous-titrage (RGST)**, créé en 1992 pour promouvoir le sous-titrage en langue française, la communauté malentendante s'est réjouie de ce projet. D'ailleurs, on envisage son adaptation à des domaines comme la justice et l'enseignement.

« Dans les cégeps, les personnes sourdes et malentendantes doivent avoir recours à un interprète ou à un volontaire qui prend des notes de cours. Avec le sous-titrage en direct, la personne pourrait, à l'aide de son portable, obtenir l'information verbale de l'enseignant d'une façon beaucoup plus précise, relire ses notes et compléter ses devoirs avec une meilleure compréhension », précise Richard McNicoll.

En direct sur la vie

À la télévision, le sous-titrage offert depuis les années 80 est plus facilement

accessible grâce au remplacement, depuis 1992, d'un onéreux décodeur par une puce intégrée dans les téléviseurs. Davantage d'émissions sont sous-titrées, mais la transcription est surtout confinée aux contenus préenregistrés, faute de temps, de ressources ou d'argent.

Le **CRTC** a imposé en 1999 une augmentation annuelle de 10 % du nombre d'émissions sous-titrées, en vue d'atteindre un taux de 90 % d'ici 2007. De plus, face à l'accroissement des interventions en direct et aux limites de la transcription par sténotypie, l'industrie de la télévision a fait appel aux technologies pour accélérer le processus.

Le **Groupe TVA** a ainsi développé, en collaboration avec le CRIM, le logiciel STDirect qui reproduit automatiquement en sous-titres les paroles prononcées sur les ondes et répétées par un locuteur. Ce logiciel utilise notamment des banques de données thématiques qui sont activées au besoin pour faciliter la reconnaissance des termes selon le contexte (sports, affaires, etc.). Progressivement utilisé pour le sous-titrage des bulletins d'information, le logiciel servira aussi à la transcription lors de l'enregistrement des émissions.

Serge Bellerose, vice-président principal, canaux spécialisés et développement des affaires, Groupe TVA, remarque que les progrès technologiques des dernières années en reconnaissance de la parole ont sans cesse augmenté la rapidité et la fiabilité du procédé informatisé. On peut même penser que le sous-titrage automatisé deviendra si performant qu'il éliminera le recours à un intermédiaire. « Nous utilisons présentement un locuteur entraîné ayant une voix neutre, dans un environnement contrôlé, pour obtenir un niveau de performance élevé. La poursuite de la recherche et du développement pourrait permettre d'éliminer le recours au locuteur dans quelques années », indique M. Bellerose.

D'ailleurs, Richard McNicoll, qui a participé au développement du logiciel STDirect, confirme que l'automatisation du sous-titrage ne peut être que bénéfique au petit écran. « Le taux de 50 % d'émissions sous-titrées est dépassé dans l'ensemble des réseaux francophones, et la proportion est de 60 % pour les chaînes conventionnelles. Si on ne tient pas compte de **Radio-Canada**, le taux serait plutôt de 45 %. Mais, avec le nouveau système, nous croyons qu'il sera possible de sous-titrer la presque totalité des émissions de télévision en 2007 ou en 2008, le temps de construire des corpus de données spécifiques aux diffuseurs. »

La voix de l'avenir

L'avenir de la reconnaissance automatisée de la parole laisse donc présager des jours meilleurs pour les malentendants.



Serge Bellerose,
Vice-président principal,
canaux spécialisés et
développement d'affaires
Groupe TVA

Selon Pierre Dumouchel, le CRIM oriente ses recherches vers la locution à distance et le traitement distribué. Les analyses des transcriptions effectuées à des endroits distants seront centralisées par la réseautique afin d'améliorer la technologie. Serge Bellerose précise, pour sa part, que les grands réseaux de télévision du Québec suivent de près le projet STDirect.