

D'une technologie à l'autre...  
plus rapidement avec le CRIM!

développement  
**logiciel**

# Sous-titres automatiques en direct



## Le CRIM aide TVA à réaliser le sous-titrage automatique de ses émissions en direct



**Le système de sous-titrage automatique qui sera développé par l'équipe de reconnaissance de la parole du CRIM aura d'importantes retombées.**

Par **Christian Aubry**

**L**e CRIM vient d'annoncer la signature d'un important contrat de recherche-développement avec le Groupe TVA inc. Il s'agit de concevoir un système de sous-titrage en temps réel permettant aux quelque 750 000 personnes sourdes et malentendantes du Québec d'accéder au contenu intégral des séquences et émissions de télévision diffusées en direct. Ce contrat aura d'importantes retombées qui pourraient profiter à plusieurs communautés francophones dans le monde.

Les recherches sur la reconnaissance automatique de la voix ont débuté il y a une quinzaine d'années, rappelle Pierre Dumouchel, Vice-président, Recherche-développement, et chercheur principal de l'équipe Reconnaissance de la parole du CRIM. «Les vocabulaires de l'époque comprenaient une dizaine de mots alors qu'ils en comptent jusqu'à plusieurs millions aujourd'hui.» Dans des conditions optimales, les logiciels actuels sont capables de transcrire efficacement 80 à 95 % des sources vocales qu'on leur soumet. Dans le cas des

reportages et des émissions en direct, cependant, la diversité des locuteurs, l'enchevêtrement des dialogues, les bruits de fond et autres impondérables sonores et linguistiques – comme les rires et les pauses d'hésitation – compliquent grandement la tâche des logiciels.

À l'heure actuelle, l'investissement en reconnaissance vocale n'est pas une priorité gouvernementale ou industrielle, constate Pierre Dumouchel. «En 1990, il y avait trois centres de recherche – à l'INRS (Institut national de recherche scientifique), à l'Université McGill et au CRIM – qui regroupaient des équipes d'une dizaine de chercheurs. Aujourd'hui, il ne reste que le CRIM», rappelait-il récemment au journal *Les Affaires*. De son côté, le secteur privé entretenait d'importants laboratoires, notamment chez Nortel et Bell Northern Research. Or, douze ans plus tard, le Québec n'aligne dans ce secteur que quelques PME réalisant des applications téléphoniques.

Selon le Regroupement québécois pour le sous-titrage (RQST) et son président,

## «Les vocabulaires de l'époque comprenaient une dizaine de mots alors qu'ils en comptent jusqu'à plusieurs millions aujourd'hui.»

M. Richard Mc Nicoll, l'information dispensée en direct par les médias électroniques est devenue aujourd'hui une ressource critique à laquelle il est très important que les malentendants puissent accéder. En 2001, le CRTC s'est rallié à cette analyse et a émis de nouvelles exigences en matière de sous-titrage. À l'instar de leurs homologues de langue anglaise, les principaux canaux de télévision francophones sont donc désormais tenus, lors du renouvellement de leur licence, de hausser le taux de sous-titrage de leurs émissions à 90 % et plus. C'est ce qui a été signifié au réseau TVA lors de son rachat par Québecor.

Pour atteindre cet objectif, il faut disposer d'une technologie de sous-titrage en temps réel afin de traiter les reportages et les segments en direct, de plus en plus fréquents. À Radio-Canada, on utilise un logiciel appartenant à IBM France et roulant sur la plateforme DOS. À l'instar de nombreux systèmes en service aux États-Unis, le sous-titrage automatique est réalisé à partir de transcriptions sténographiques manuelles. Les notes de sténo génèrent cependant de nombreuses erreurs, sans parler du retard à l'affichage et de la difficulté à faire évoluer un logiciel propriétaire relativement ancien.

La technologie mise en oeuvre par le CRIM est radicalement différente. La source du sous-titrage sera la source sonore elle-même. Dans un premier temps, on aura recours à un locuteur dédié – un «perroquet», en jargon scientifique – qui répétera, en studio et en temps réel, les propos à retranscrire. Grâce à cet intermédiaire intelligent et coopératif, il sera possible d'atteindre les conditions optimales garantissant un taux de reconnaissance et de transcription satisfaisant dès que le système sera opérationnel, soit en février 2004. Par la suite, il faudra réaliser des progrès technologiques plus significatifs avant d'être en mesure de se passer des «locuteurs» professionnels.

### Pierre DUMOUCHEL

*Vice-président, Recherche-développement, et  
chercheur principal de l'équipe  
Reconnaissance de la parole du CRIM*



Réalisé par le CRIM, en collaboration avec le RQST, ce contrat coûtera au Groupe TVA quelque 500 000 dollars, auxquels s'ajouteront d'importants frais en personnel, en formation, en aménagements de locaux et en administration. Cela reste malgré tout un bon investissement, explique Linda Malenfant, Directrice, Support à l'information, du télédiffuseur : «Sans ce projet, nous devrions dépenser des dizaines de milliers de dollars par année en personnel afin d'atteindre l'objectif fixé par des moyens classiques. Nous préférons investir dans le développement d'un outil qui nous permettra de mettre en ondes sans recorriger le texte. L'avantage déterminant, ce sera le gain en productivité.»

Selon Pierre Dumouchel, «la puissance des machines est un paramètre significatif pour le temps réel, mais l'intelligence est plus importante encore. À chaque fraction de seconde, l'ordinateur doit se poser des centaines de questions : s'agit-il d'un mot nouveau, ou du milieu d'un mot, etc. Il y a des milliers d'hypothèses et beaucoup de redondance.» L'intelligence logicielle développée par le CRIM permet de simplifier l'analyse. L'identification d'une branche terminologique fondée sur le préfixe «télé», par exemple, restreindra la recherche aux seuls mots comportant cette racine. De la même manière, tous les mots finissant par le suffixe «tion» seront regroupés, ce qui aura pour résultat de compresser le graphe de recherche et d'accélérer le processus.

La réalisation de banques de corpus de données audio-linguistiques spécialisés (pour les

sports, les discours culturels, juridiques, économiques, etc.) constitue un axe de développement parallèle. Ces banques permettent aux ordinateurs de disposer du vocabulaire adéquat et de mieux circonscrire les propos qu'ils doivent identifier. On peut ainsi éviter, par exemple, que la phrase «Pour les femmes, le port du chapeau est obligatoire» soit transcrite à l'écran par «Poulet femmes le porc du chat peau est obligatoire», comme cela s'est produit récemment à la télévision! Craignant que le système de sous-titrage par locuteur soit un peu lourd à gérer, Richard Mc Nicoll presse donc les gouvernements à «investir les sommes adéquates afin que ces corpus existent et soient mis à la disposition des chercheurs de tous les milieux.»

À terme, M. Mc Nicoll croit que le développement entamé par TVA aura des retombées positives pour plusieurs communautés de personnes handicapées, comme les aphasiques, les dysphasiques et certains cas de paralysie. «Un des champs qui nous vient à l'esprit, également, c'est celui de la robotique», ajoute-t-il. Selon Mme Malenfant, il pourrait également y avoir des retombées intéressantes dans l'industrie du doublage de films, sans parler des autres sociétés francophones de radiodiffusion qui devraient au premier chef se montrer intéressées par cette technologie. Pour M. Dumouchel, enfin, le vieillissement de la population risque d'entraîner un accroissement des pathologies auditives, rendant les applications de reconnaissance de la parole de plus en plus indispensables. ☐