



Interface n° e-96 Septembre 2004

La reconnaissance vocale

L'évolution de l'informatique permet de nos jours d'oublier le clavier et la souris et pourtant de pouvoir encoder du texte ou naviguer dans les menus et les boîtes de dialogues, ceci grâce à un logiciel de reconnaissance vocale. Il s'agit en quelque sorte d'une dictée: l'utilisateur parle à son ordinateur et celui-ci écrit. Le programme essaye de reconnaître les mots prononcés pour les retranscrire dans un traitement de texte. Nous avons reçu dernièrement un lot de documents qui ont dû être traités de cette manière. Ce sont les ratés du système qui nous ont permis de faire cette déduction!

Dans quelque 200 pages, nous avons repéré plusieurs passages isolés (une à deux pages chaque fois) agrammaticaux, dépourvus de sens et sans ponctuation. En voici quelques extraits:

- Yi Zhu vis la face de Dieu panda un instar (comprenez: Jésus vit la face de Dieu pendant un instant)
- Tout l'ai jour on ses machines... (comprenez: Tous les jours, on s'imagine...)
- Hystérectomie (comprenez: iste rex gloriae) – etc.! Celui qui cherche à deviner ce qui se cache derrière ces drôles de suites de mots a à jouer avec l'homophonie globale de la "phrase".

Plusieurs facteurs influent sur les performances de la technologie de la reconnaissance vocale. Deux grands systèmes coexistent: la reconnaissance vocale en discontinu et en continu. Le système discontinu est le plus ancien et le plus performant. Il demande à l'utilisateur de laisser un temps mort entre chaque mot prononcé pour permettre à la machine de le saisir. Par contre, dans le système en continu, où le lecteur parle de façon naturelle, la reconnaissance s'effectue en fin de phrase. C'est un procédé délicat pour une machine qui n'est pas intelligente. Ce n'est déjà pas évident pour elle de trouver la fin et le début des mots [exemples: donner de quoi a saisi et sa fin (= donner de quoi rassasier sa faim) – il de mœurs dans la vérité (= il demeure dans la vérité)]. Ensuite la production de chaque son est affectée par la production des phonèmes avoisinants [exemples: car réel ne disait rien (= car elle ne disait rien) – il faut que vous soyez Savard pour... (= il faut que vous soyez avare pour...)]. Une élocution rapide complique aussi beaucoup.

Un deuxième critère est le degré de dépendance. On parle d'un système dépendant s'il est adapté ou adaptable (après un "test d'entraînement") à un locuteur déterminé ou à un groupe défini de locuteurs, et d'un système indépendant s'il est adapté à toute personne. Il faut savoir que l'étape initiale d'un programme de reconnaissance vocale consiste à modéliser le spectre vocal de l'utilisateur (avec sa prononciation particulière, son intonation, la situation dans laquelle il parle). Une hypothèse que l'on pourrait peut-être formuler concernant les passages problématiques rencontrés, est qu'il s'agirait d'un autre locuteur qui aurait pris le relais pendant la dictée, sans que la machine ait son empreinte vocale.

Un troisième paramètre est la taille et la complexité du vocabulaire. En effet, certaines applications ont seulement besoin de quelques mots (rien que des nombres, par exemple), d'autres, au contraire, nécessitent un très large vocabulaire. Des dictionnaires de langue, plus ou moins étendus, complètent l'automatisme du processus. S'il est question d'une terminologie très bien définie, un dictionnaire spécialisé sera nécessaire et suffisant. Dans ce cas, on peut s'attendre à obtenir des résultats plus précis et exacts, car un tel contexte linguistique appelle un tel mot; la machine prévoit les occurrences probables. Mais quand il

s'agit de littérature, voire de poésie... le problème de l'univocité du langage se pose avec acuité! La technologie de la traduction automatique rencontre les mêmes impasses... (Autre remarque: dans nos fameux extraits, nous avons repéré quelques anachronismes lexicaux pour la langue d'un auteur du 19e siècle: rock, great, Senghor, etc.!).

Un dernier facteur est le coût et la méthode d'implémentation. Par exemple, le système de reconnaissance vocale en continu exige un PC "musclé" (puissant, pour calculer les combinaisons possibles des mots, et doté d'un volume important de mémoire).

Les technologies de la reconnaissance vocale se sont largement développées au cours de la dernière décennie, compte tenu du large éventail d'applications potentielles. Toutefois, ce processus ne semble pas être actuellement tout à fait au point. Le sera-t-il jamais?

Anne-Françoise Denamur

