

Parler n'est pas suffisant

La communication naturelle ne se limite pas aux échanges oraux. De ce fait, une interaction homme-machine n'est idéale que si la communication est «multimodale», c'est-à-dire si elle intègre la parole, mais aussi les gestes de désignation et la perception visuelle. Lorsque nous parlons, nous faisons souvent référence à des objets de notre environnement, que nous voyons et désignons, par des déterminants ou pronoms particuliers, tels «une», «la», «cette», «celle-ci»...

Naturelle chez les êtres humains, cette multimodalité pose à la machine des problèmes d'interprétation et d'influence réciproque entre les modes d'interaction. Par exemple, avec une interaction associée à deux désignations, telle «place cette figure au-dessus de ce texte», la parole, le geste et la perception ne peuvent pas s'interpréter indépendamment.

Depuis dix ans, notre équipe, composée d'informaticiens et de linguistes, étudie la mise en place d'une telle communication : à partir d'études linguistiques, cognitives et informatiques, nous élaborons tout d'abord des modèles informatiques de représentation des connaissances, puis des algorithmes pour traiter un problème particulier, dont nous testons ensuite les capacités de reconnaissance, de compréhension et d'interprétation sur des applications concrètes. Quand l'algorithme est performant, nous l'intégrons enfin à un logiciel ou à une interface.

Ainsi, un de nos modèles linguistiques intègre et comprend les hésitations, les reprises et les corrections, nombreuses à l'oral et qui, bien que diverses, respectent une structure prévisible. Un autre modèle est spécialisé pour interpréter les gestes de désignation naturelle de l'utilisateur, et un troisième modèle, dit de représentation mentale, unifie l'ensemble des informations connues sur chaque composant d'une tâche. Un autre modèle tient compte des présupposés inclus dans les énoncés d'un utilisateur afin de mieux adapter les réponses de la machine et de gérer le dialogue.

Par ailleurs, nous cherchons des algorithmes qui exploitent les connaissances de ces modèles pour comprendre les énoncés multimodaux. Ainsi un de nos algorithmes interprète les différentes références incluses dans une tâche, notamment quand celle-ci est visualisée sur un écran, et intègre le fonctionnement des déterminants en français. S'appuyant sur un pavage de l'espace cognitif, par exemple, de l'écran, l'algorithme lève les ambiguïtés d'énoncés, tels «la figure», «cette figure», «une de ces figures». Nos algorithmes sont écrits en Java, le langage en vigueur sur le réseau *Internet*, assurant ainsi leur adéquation aux divers systèmes informatiques.

Enfin, nous avons proposé une nouvelle architecture, modulable et répartie, pour les systèmes de dialogues multimodaux, qui s'adapte d'une part au modèle de l'application à gérer, c'est-à-dire à la définition des objets à traiter, à leurs propriétés et aux fonctions qui leur sont applicables. D'autre part, l'architecture s'adapte aux spécificités lexicale, syntaxique, sémantique et d'usage des différents modes de communication ; ainsi, pour le langage, nous traitons la langue soutenue, technique, écrite et orale.

Bien que nos recherches soient encore fondamentales, des applications industrielles sont déjà en cours. Ainsi, avec la Société *Alcatel*, nous avons mis au point une interface de communication, répondant aux gestes et à la voix, qui équipe un terminal de télécommunication composé d'un écran tactile et d'un combiné téléphonique. Nous mettons également au point un environnement multimodal de consultation d'annuaires complexes pour ce type de terminal de télécommunication qui pourrait bientôt apparaître sur le marché.

À l'avenir, seule une communication multimodale nous permettra d'accéder de façon naturelle aux nombreuses applications interactives qu'offrent aujourd'hui le réseau *Internet* et la télévision numérique.

Jean-Marie PIERREL,
Université de Nancy et Équipe
«Langue et Dialogue» du LORIA

transférés directement du *Net 21* sur les *Handy 21* et *Enviro 21*, en fonction des demandes, des erreurs et des mises à jour des utilisateurs. Les utilisateurs adapteront les machines qui les entourent à leurs besoins et à leurs habitudes.

Une conviction et un souhait

Oxygen nécessite des ordinateurs portables, des ordinateurs muraux et des ordinateurs embarqués, un réseau innovant, des systèmes de compréhension de la parole et de synthèse vocale, l'accès aux connaissances, la collaboration, l'automatisation et la personnalisation. *Oxygen* devra sa puissance à l'ensemble de ses éléments, qui sont tous tournés vers l'utilisateur. La communication avec les ordinateurs a commencé avec des nombres binaires. Puis on a utilisé des textes et, récemment, des icônes ; il est temps d'améliorer encore les interfaces homme-machine.

Je crois que l'informatique du futur dépendra essentiellement des échanges vocaux (ou à l'aide d'autres capacités sensorielles), de l'accès aux connaissances, de l'automatisation, de la collaboration et de la personnalisation. Ces cinq aspects, à mon avis incontournables, sont le volant, l'accélérateur et les freins que l'informatique utilisera pour nous en donner plus.

Je propose que tous ceux qui souhaiteront tirer profit du nouveau monde de l'information exploitent les potentialités du système *Oxygen*. Les médecins, par exemple, seraient plus efficaces s'ils disposaient à tout instant des données de *Medline* (qui regroupent des articles médicaux de très nombreuses revues), des dossiers médicaux de leurs patients et des avis de collègues spécialisés.

L'informatique, avec *Oxygen* et d'autres systèmes de même nature, doit maintenant nous libérer des détails techniques que nous supportons depuis 40 ans et nous faire entrer dans une nouvelle ère : celle des ordinateurs au service de l'utilisateur. Après la charrue, le moteur et l'ordinateur, la quatrième révolution concerne la ressource terrestre la plus précieuse : nous-mêmes.

Michael DERTOUZOS est directeur du Laboratoire d'informatique de l'Institut de technologie du Massachusetts.