

de connaissances pour répondre à la requête d'un ou de plusieurs utilisateurs simultanément. On accède à *Galaxy* par téléphone ou par ordinateur, avec une connexion au réseau *Internet*, qui transmet les données.

Galaxy assure cinq fonctions : la reconnaissance de la parole, la compréhension du langage, l'extraction de l'information, l'élaboration du langage et la synthèse vocale. Quand on pose une question, un serveur nommé *Summit* analyse ce qu'il entend à l'aide d'une bibliothèque de phonèmes (les plus petites unités sonores qui constituent les mots, dans une langue), puis il dresse une liste ordonnée de phrases possibles. Un deuxième serveur, nommé *Tina*, applique les règles grammaticales, décompose la phrase en éléments (sujet, verbe, complément...) et traduit alors l'énoncé le plus plausible en un bloc sémantique, c'est-à-dire une série de commandes que la machine comprend. Par exemple, *Tina* reformulera la question : « Où se trouve le musée d'Orsay ? » en la commande suivante : « Localiser le musée nommé musée d'Orsay. »

Ensuite, un troisième serveur, *Genesis*, convertit le bloc sémantique en une requête adaptée au format de la banque de données la plus pertinente. *Tina* réorganise alors l'information extraite en un nouveau bloc sémantique, que *Genesis* convertit en une phrase dans la langue de l'utilisateur : « Le musée d'Orsay se trouve au numéro 1 de la rue de la Légion d'honneur, à Paris, dans le VII^e arrondissement. » Enfin, un synthétiseur vocal dit la phrase.

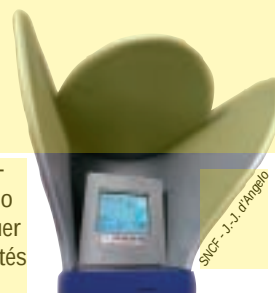
Pour passer d'un système à l'autre, l'utilisateur n'a qu'à demander : « Je veux parler à *Jupiter* » ou : « Passe-moi *Voyager*. » Depuis mai 1997, *Jupiter* a répondu à plus 30 000 appels, et le taux de compréhension des requêtes d'utilisateurs novices atteint 80 pour cent.

Le projet MASK

Dans les halls de gare, des billetteries automatiques à interface tactile sont destinées à faciliter l'accès de la clientèle au système de distribution. Malheureusement, la rigidité et le manque de convivialité de ce système n'ont pas séduit les voyageurs. Depuis 1994, dans le cadre d'un programme de recherche européen, en collaboration avec le CNRS, l'Université de Londres et les Sociétés *Mors* et *Signes particuliers*, la SNCF met au point une borne d'information et de distribution qui, à terme, remplacera les billetteries actuelles.

Il s'agit d'un kiosque de réservation et d'achats de billets de train à interface naturelle multimodale : la machine répond aux paroles ou au toucher de l'utilisateur par une voix de synthèse, des séquences vidéo ou des textes. Le kiosque, nommé *MASK*, est composé de deux ordinateurs : une station de travail *Silicon Graphics* qui gère les différents modules, ou logiciels, du système, et un ordinateur personnel classique, qui se charge de l'interface graphique. Lorsqu'un utilisateur énonce une requête, un premier module de reconnaissance de la parole, qui intègre les hésitations, traduit les signaux acoustiques en signaux numériques. Un deuxième module extrait de ces signaux les informations requises (date de départ, destination, gare de départ...) qu'un troisième module, le gestionnaire de dialogue, recherche dans une banque d'information, nommée *Riho*. Le lexique répertorie aujourd'hui plus de 1 500 mots, dont les noms de 600 villes françaises. Quand des informations sont encore nécessaires, un module de génération de phrase formule une question qu'un synthétiseur vocal prononce. Lorsque la requête est complète, une liste de trains est proposée sur l'écran, et le voyageur n'a plus qu'à choisir. L'achat d'un billet peut durer moins d'une minute.

Le kiosque a été testé pendant un mois à la gare Saint-Lazare par des clients volontaires qui suivaient un scénario préétabli. On analyse ces essais préliminaires afin d'évaluer la machine en conditions réelles, et d'améliorer ses capacités de compréhension et de dialogue.

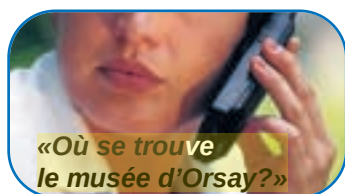


Nous enregistrons les appels, puis nous les analysons, afin d'améliorer les performances du système.

Nos systèmes de reconnaissance de la parole seraient parfaitement adaptés aux appareils portatifs *Handy 21* du projet *Oxygen* : le clavier, notamment, pourrait être supprimé, et le langage parlé permettrait aux utilisateurs de communiquer plus efficacement avec leur appareil. Ainsi, un homme d'affaires en déplacement pourrait demander à son ordinateur : « Dis-moi quand l'action Crédit lyonnais dépassera 210 francs. » La machine deviendra ainsi un assistant qui accomplit des tâches variées à partir d'un minimum d'instructions.

Nous devons encore progresser avant d'atteindre ce but. Notamment, nous devons réaliser des programmes capables de traiter de l'information dans des domaines complexes, nous devons augmenter le nombre de langues comprises par la machine et nous devons obtenir des systèmes qui réunissent des données variées sans y être explicitement invités. L'ordinateur ne doit pas seulement comprendre ce que l'on dit, mais plutôt ce que l'on veut dire.

Victor ZUE est directeur adjoint du Laboratoire d'informatique de l'Institut de technologie du Massachusetts.

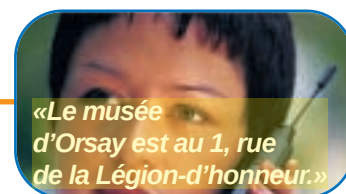


SUMMIT
comprend
la phrase

TINA
formule
la demande

GENESIS
recherche
l'information

TINA
organise
la réponse



LES APPLICATIONS À RECONNAISSANCE VOCALE utiliseront l'architecture *Galaxy* pour répondre à la question d'un utilisateur. Le système *Summit* établit une liste d'interprétations de la phrase, puis le

système *Tina* traduit la proposition la plus plausible en instructions qui permettent au système *Genesis* d'extraire l'information d'une banque de données. Enfin, le système *Tina* formule une réponse.