

## **Le laboratoire de traitement du langage et de la parole de l'Université de Neuchâtel: recherche fondamentale et appliquée**

Créé en 1987, le laboratoire de traitement du langage et de la parole est rattaché à l'Institut de linguistique de l'Université de Neuchâtel et poursuit deux types de recherche. En premier lieu, il étudie le traitement naturel de la parole et du langage (également appelé psycholinguistique expérimentale), à savoir la perception, la compréhension et la production chez l'être humain. En deuxième lieu, et en collaboration étroite avec certaines entreprises de technologie avancée, il entreprend des travaux dans le domaine du traitement automatique de la parole et du langage (linguistique-informatique), soit la synthèse et la reconnaissance de la parole, l'analyse automatique du langage, la correction de textes et la traduction assistée par ordinateurs.

### *1) Description du laboratoire*

Le personnel du laboratoire est composé d'un directeur, de deux chercheurs rétribués sur des fonds de recherche privés et de deux assistants d'université (postes partiels). Les compétences de ce personnel<sup>1</sup> couvrent les domaines suivants: informatique, acoustique, phonétique, psycholinguistique expérimentale, tests et évaluation, linguistique générale, statistique. Les trois langues de travail sont le français, l'anglais et l'allemand.

Le laboratoire occupe trois pièces dans le bâtiment principal de l'université et est équipé du matériel suivant: ordinateurs divers (PC AT, Macintosh, Rainbow, liaison avec les Vax du Centre de calcul), instruments d'analyse de la parole (sonographe, visipitch, analyseur de spectre), systèmes d'analyse informatisés (MacAdios, MicroSpeech Lab), synthétiseurs de parole (Infovox, DecTalk, Synthé), reconnaisseurs de parole (TI-Speech, RDP8a), logiciels de traduction assistée par ordinateurs (TSS de ALP Systems), langages de programmation (LISP, PROLOG) et programmes de statistique (Abstat, StatView).

Le laboratoire est subventionné par l'Université de Neuchâtel, par des fondations publiques de recherche et également par le privé. En effet, plus de la moitié de son matériel provient de dons en provenance d'entreprises en Suisse et à l'étranger.

<sup>1</sup> Outre le directeur, les membres du laboratoire de 1987 et 1989 ont été: JEAN-YVES DOMMERGUES (Maître de conférences à l'Université de Paris VIII), LYSIANE GROSJEAN, NATALIE KÜBLER et ALAIN MATTHEY.

## *II) Recherche fondamentale*

### *a) La psycholinguistique expérimentale*

Nos études portent sur le traitement naturel de la parole et font suite à celles que nous avons entamées aux Etats-Unis. Notre démarche est celle de la psycholinguistique expérimentale, discipline qui a recours en priorité à l'expérience scientifique afin de découvrir les processus mentaux sous-jacents au traitement du langage. Cette préférence est dictée par la nécessité. Il est fort difficile, en effet, d'étudier la perception, la compréhension et la production du langage à l'aide d'outils uniquement descriptifs, introspectifs ou théoriques. Le psycholinguiste est condamné, en conséquence, à conduire des expériences. Pour ce faire, il suivra un certain nombre d'étapes. Il partira d'un modèle de traitement et cherchera à le modifier ou à le compléter. Il formulera donc certaines hypothèses qu'il tentera de confirmer ou d'infirmer à l'aide de l'expérimentation. Il choisira les matériaux linguistiques qui seront soumis à des sujets et optera pour un plan d'expérience dans lequel seront définies les variables dépendantes, indépendantes et de contrôle ainsi que l'ordre de présentation des matériaux. Le chercheur définira alors une tâche expérimentale et sélectionnera un groupe de sujets représentatifs de la population étudiée. L'expérience se déroulera d'ordinaire dans un laboratoire équipé des appareils nécessaires (ordinateurs, instruments de mesures, magnétophones, etc.) et les données recueillies seront ensuite soumises à des traitements d'ordre statistique. Le chercheur sera alors en mesure de vérifier si les hypothèses émises au départ ont été confirmées ou infirmées et, selon la réponse, il proposera des modifications plus ou moins importantes du modèle retenu au départ.

Il existe une étape particulièrement importante dans cette procédure. Il s'agit du choix de la tâche que l'on demandera aux sujets d'effectuer. Celle-ci, en effet, permettra d'étudier indirectement les opérations sous-jacentes qui ont normalement lieu lors du traitement du langage, opérations qui ne sont pas directement observables par le chercheur. Comme toute science expérimentale, la psycholinguistique est tributaire de la qualité et du bien fondé des procédures qu'elle utilise. Ces dernières doivent pouvoir «ouvrir une fenêtre» sur le fonctionnement interne du système psycholinguistique et permettre au chercheur d'obtenir le plus d'informations possible en évitant d'introduire des interférences avec le processus étudié. Ce dernier point est crucial. En effet, nombre d'études révèlent davantage les opérations inhérentes à la tâche elle-même que celles qui sous-tendent le mécanisme qu'on est censé étudier.

La tâche du chercheur en psycholinguistique consiste donc à contribuer à la découverte des processus mentaux impliqués dans la perception, la

compréhension et la production d'un énoncé linguistique. Pour arriver à cette fin, il ne lui suffit pas de faire passer des expériences et d'obtenir des données. Il doit également se servir de ses résultats pour confirmer ou infirmer un modèle ou une de ses composantes. C'est au cours des années soixante-dix que des modèles spécifiquement psycholinguistiques furent proposés. Ceux-ci mettent en exergue le traitement en temps réel, la construction active de la représentation interne, l'apport spécifique de chacun des modules d'analyse qui fonctionnent en parallèle, l'intégration de l'information obtenue par ceux-ci, le rôle du traitement ascendant et descendant ainsi que de l'anticipation. De nouveaux arguments théoriques viennent justifier ces modèles et sont, à leur tour, renforcés par les résultats obtenus grâce à de nouvelles techniques expérimentales.

### *b) La psycholinguistique du bilinguisme*

Après un certain nombre de travaux sur les variables temporelles, les unités de production, l'accès au lexique, la perception de la parole et de la prosodie, et le traitement de la langue des signes (voir, par exemple, GROSJEAN, 1979; GROSJEAN, 1980; GROSJEAN et DESCHAMPS, 1975; GROSJEAN et DOMERGUES, 1983), nos efforts portent depuis quelques années sur la psycholinguistique du bilinguisme. Les premières études dans ce domaine, effectuées en grande partie au Canada, ont cherché, sans beaucoup de succès, à montrer la réalité psychologique de la distinction généralement admise entre bilingues coordonnés, composés et subordonnés. D'autres ont examiné l'organisation des lexiques internes chez les bilingues, et d'autres encore ont comparé le comportement psycholinguistique du bilingue et du monolingue.

Ces études ont souvent été conduites dans une optique que l'on pourrait qualifier de «monolingue» ou «fractionnelle», celle qui veut que le bilingue possède deux compétences linguistiques bien spécifiques et isolables. Selon ce point de vue, le bilingue serait, en quelque sorte, deux monolingues réunis en une seule et même personne. Ses compétences linguistiques, ainsi que les opérations de traitement du langage qu'il met en œuvre, seraient donc comparables à celles du monolingue.

Depuis plusieurs années, nous défendons une autre perspective du bilinguisme, une optique que nous intituleons «bilingue» ou «holistique» (GROSJEAN, 1982; GROSJEAN, 1984). Le bilingue, selon cette théorie, est un tout indissociable; il manifeste une compétence linguistique globale, unique et spécifique, qui est difficilement décomposable en deux ou plusieurs compétences monolingues. La coexistence et l'interaction de deux langues chez le bilingue a créé un ensemble linguistique différent de celui du mono-

lingue, ensemble qui devra donc être étudié en tant que tel, aussi bien par le linguiste que par le psycholinguiste. Un exemple de cette particularité est ce qu'on appelle le parler bilingue, à savoir l'utilisation momentanée d'une deuxième langue (sous forme d'alternances codiques et d'emprunts) lorsque le bilingue communique avec un autre bilingue dans une langue donnée. Un programme de recherche destiné à étudier la psycholinguistique du parler bilingue est en cours depuis plusieurs années. Nous essayons de répondre aux questions suivantes: Lors de la production d'une alternance codique, le locuteur passe-t-il complètement d'un système phonétique à un autre? Ce passage se fait-il aussi bien au niveau segmental qu'au niveau prosodique? Qu'en est-il de l'emprunt idiosyncrasique? Et l'auditeur, comment traite-t-il un énoncé mixte? Comment accomplit-il le décodage acoustique et phonétique? Quel est le cheminement qui permet l'accès aux deux lexiques internes?

Dans une étude récente (GROSJEAN, 1988), nous avons tenté d'isoler les variables qui jouent un rôle dans l'accès au lexique lorsque l'énoncé est mixte. Des mots anglais, produits sous forme d'alternances codiques et d'emprunts dans des phrases françaises, étaient présentés à des sujets bilingues français-anglais à l'aide de la méthode de reconnaissance de mots appelée «gating». Les résultats obtenus ont montré l'importance des facteurs suivants: la configuration phonotactique du mot (est-elle indiscutablement spécifique à une langue ou se retrouve-t-elle dans les deux?); la présence d'un homophone dans l'autre langue; la langue de l'énoncé qui précède le mot, et enfin, le statut phonétique de l'item en question – est-il prononcé dans la langue de base ou dans l'autre langue? Un modèle qui se fonde sur l'approche connexionniste a été développé afin de rendre compte des résultats obtenus.

A la longue, nous devrions pouvoir proposer, et par la suite affiner, un modèle de traitement du parler bilingue, modèle qui comportera des aspects communs à tout modèle de traitement et d'autres qui seront spécifiques aux interlocuteurs bilingues et à leur manière de communiquer.

### *III) Recherche appliquée*

Une des tâches confiées au laboratoire lors de sa création était de se mettre à la disposition de groupes extérieurs (centres de recherche, entreprises, etc.) afin de collaborer à des projets communs. Ce mandat a été pris au sérieux et une vingtaine de travaux ont été menés à bien dans les deux premières années. Parmi ceux-ci nous trouvons:

#### *a) L'élaboration d'un continuum acoustique*

A la demande du Centre de neuropsychologie du CHUV, qui cherchait à compléter ses batteries de tests de perception de la parole, le laboratoire a élaboré un continuum acoustique «camp-gant» pour le paradigme de la perception catégorielle. A partir des phrases «Montrez le camp» et «Montrez le gant», et en utilisant une méthode d'hybridation, nous avons construit 10 phrases intermédiaires en remplaçant par étapes le silence qui précède le /k/ de «camp» par le prévoisement du /g/ de «gant», tout en préservant la proportionnalité relative de leur durée. Nous avons ensuite conduit une expérience pilote afin de vérifier que les deux extrémités du continuum étaient perçues de façon non ambiguë, qu'il existait bien une frontière perceptive à l'intérieur du continuum et que celle-ci était nette. Les résultats ont révélé la manifestation de la perception catégorielle chez chacun des sujets.

Depuis un an le continuum sert à tester les capacités d'identification et de discrimination de sujets normaux, de patients aphasiques et d'enfants dyslexiques.

#### *b) L'évaluation d'un système de reconnaissance de parole*

A la demande de Systèmes G (Lannion, France), le laboratoire a procédé à l'évaluation du nouveau reconnaisseur de parole «RDP8-A», système multilocuteur ayant une capacité de 8 ordres. Ce travail consistait principalement à mesurer les performances de l'appareil dans des conditions d'utilisation diverses (environnement calme et bruité, micro à différentes distances, etc.), avec des voix masculines et féminines et en présentant des mots «légaux» (les ordres censés être reconnus) et des mots «illégaux» (mots proches et mots différents des mots légaux).

Nous avons pu montrer que le système donne un taux global de reconnaissance élevé (86% en condition normale), qu'il reconnaît presque aussi bien les voix féminines que les voix masculines et qu'il résiste plutôt bien aux diverses conditions de l'évaluation sauf lorsqu'il est testé dans le bruit ou que le micro se trouve à une trop grande distance du locuteur.

Aujourd'hui, le reconnaisseur de Systèmes G est intégré dans de nombreuses applications nécessitant l'entrée vocale.

#### *c) Travaux autour de la téléthèse de communication «Hector»*

Depuis l'établissement du laboratoire en 1987, nous avons entrepris un certain nombre de projets pour la Fondation suisse pour les téléthèses, la majorité ayant trait à la téléthèse de communication «Hector». Dans une étude

récente, nous faisons le point sur la version courante d'Hector et examinons l'apport potentiel de l'intelligence artificielle et du traitement automatique du langage naturel à une nouvelle version. Outre l'utilisation d'un ordinateur plus puissant, une synthèse améliorée et une meilleure ergonomie (notamment l'écran), nous proposons que le nouveau logiciel prenne certaines responsabilités dans l'élaboration et la génération des énoncés. Il entrerait en interaction avec l'utilisateur et lui poserait des questions de manière conviviale sur le contenu de l'énoncé (sémantique et pragmatique) et sa forme (syntaxe, lexique, morphologie). Cela permettrait au programme d'organiser en partie l'énoncé et lui donnerait les informations nécessaires pour sa construction. De plus, le logiciel serait capable d'interpréter les segments ou les mots incomplets qui lui seraient donnés. Il devrait pouvoir analyser et corriger l'orthographe de ce qui est écrit et traiter correctement une morphologie ou une syntaxe partiellement défectueuse.

La nouvelle version d'Hector (la troisième) est actuellement en cours de développement dans les laboratoires de la Fondation suisse pour les téléthèses.

#### d) Contribution aux logiciels d'aide à la rédaction

Depuis un an nous collaborons à la réalisation de logiciels d'aide à la rédaction produits par Lexpertise Linguistic Software de Vaumarcus. Ces outils sont utilisés en parallèle avec un traitement de texte pour vérifier certains aspects de la prose: orthographe, ponctuation, accord, choix approprié du vocabulaire, etc. Les produits sont de deux sortes: ceux destinés à des monolingues et ceux élaborés spécifiquement pour des bilingues qui désirent corriger un texte rédigé dans leur deuxième ou troisième langue. Outre un travail d'évaluation des logiciels offerts sur le marché, à la fois aux États-Unis et en France, notre contribution a surtout consisté à créer des bases de données d'erreurs potentielles pour les logiciels bilingues. Parmi les catégories sur lesquelles nous avons travaillé, nous trouvons: les confusions (ex. «parc» français qui peut se traduire par «park, pen, enclosure, fleet» en anglais), les faux amis («experience», «training»), les mots qui sont suivis d'un verbe au singulier («news») ou au pluriel («cattle»), l'utilisation des prépositions (\*make on, \*care over), les abréviations («a.b.s.» se traduit «c/o»), les acronymes (AELE se traduit EFTA), etc.

Les premiers produits utilisent une approche dite de la mise en correspondance («pattern matching») mais ceux de la deuxième génération feront une analyse morphologique et syntaxique de la phrase et pourront proposer des corrections à ces niveaux-là également.

#### Conclusion

Dans les années qui viennent, les chercheurs du laboratoire poursuivront leurs travaux en traitement naturel, à savoir sur la perception, la compréhension et la production du langage et de la parole chez l'être humain, qu'il soit monolingue ou bilingue. Dans le domaine du traitement automatique, ils continueront à accepter des mandats en provenance des industries de langue, et entreprendront également des projets de recherche fondamentale, notamment dans le domaine des analyseurs morphologiques et syntaxiques d'une langue seconde. Ils contribueront ainsi aux sciences du langage à l'Université de Neuchâtel et au développement des entreprises de traitement automatique implantées dans le canton.

Université de Neuchâtel  
Laboratoire de traitement  
du langage et de la parole  
CH-2000 Neuchâtel

FRANÇOIS GROSJEAN

#### Références bibliographiques

- GROSJEAN, François (1979): «La psycholinguistique de la langues des signes». In: *Langages*, 56, 35-57.
- GROSJEAN, François (1980): «Spoken word recognition processes and the gating paradigm». In: *Perception and Psychophysics*, 28, 267-283.
- GROSJEAN, François (1982): *Life with two languages: An introduction to bilingualism*, Cambridge (Mass.), Harvard University Press, 370 p., ISBN 0-674-53091-8.
- GROSJEAN, François (1984): «Le bilinguisme: vivre avec deux langues». In: *Travaux neuchâtois de linguistique (TRANEL)*, 7, 15-42.
- GROSJEAN, François (1988): «Exploring the recognition of guest words in bilingual speech». In: *Language and Cognitive Processes*, 3, 233-2174.
- GROSJEAN, François et DESCHAMPS, Alain (1975): «Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation.» In: *Phonetica*, 31, 144-184.
- GROSJEAN, François et DOMMERGUES, Jean-Yves (1983): «Les structures de performances en psycholinguistique». In: *L'année Psychologique*, 83, 411-458.