

Perception et traitement de la parole

Noël Nguyen

*Laboratoire Parole et Langage, Aix-Marseille Université et CNRS
Institute for Language, Communication and the Brain, Aix-Marseille Université*

In : *Un Monde Commun — Les Savoirs des Sciences Humaines et Sociales* (CNRS Éditions, Paris), pp. 284-307

Dans la communication parlée, chaque interlocuteur déploie un large ensemble de processus de traitement afin de comprendre ce qui lui est dit à partir de la forme matérielle du message reçu. Dès lors que cet échange a lieu dans une langue que chacun maîtrise bien, le système de traitement se caractérise par une rapidité d'exécution, une efficacité, et une robustesse remarquables au regard de la vitesse à laquelle les mots se succèdent dans la parole, de la grande variabilité intra- et inter-individuelle dans la façon de produire ces mots, et du fait que le flux de parole que nous traitons est entremêlé à d'autres voix et aux sons les plus divers dans la plupart des situations de communication. Les recherches menées dans ce domaine visent à nous permettre de comprendre la nature de ces processus de traitement à l'interface entre langage, cognition et cerveau.

La place dévolue à notre « monde commun » dans ces recherches a longtemps été limitée. Cela tient à ce qu'elles ont d'abord été fortement influencées par un modèle de la communication selon lequel le traitement de la parole s'assimile au décodage par l'auditeur d'un signal d'entrée, abstraction faite des interactions sociales à l'intérieur desquelles ces opérations s'accomplissent. Ce modèle a prévalu pour des raisons à la fois méthodologiques et théoriques. Sur le plan méthodologique, et dans un cadre expérimental, il est plus facile de contrôler les facteurs pouvant exercer une influence sur les réponses des auditeurs à des stimuli langagiers, lorsque ces auditeurs sont placés en situation de réaliser individuellement la tâche, en l'absence d'une interaction avec autrui (qu'il s'agisse d'identifier des phonèmes, de dire si le stimulus perçu est un mot ou un non-mot, ou de beaucoup d'autres tâches possibles). Sur le plan théorique, il a longtemps été considéré que, dans la mise en correspondance entre parole et sens, le système de traitement faisait abstraction des propriétés liées à l'identité individuelle et sociale du locuteur dans la parole, en vertu de l'hypothèse selon laquelle ces propriétés ne jouaient pas de rôle dans l'accès au sens par l'auditeur.

Dans une large partie des études sur la perception de la parole, il s'est agi d'examiner avec précision la contribution d'un ou de plusieurs paramètres acoustiques aux distinctions perçues par l'auditeur entre phonèmes, et ces études peuvent apparaître comme étant plus proches de la psychophysique que des SHS. Or, les frontières qui s'établissent entre phonèmes dans l'espace perceptif sont dans une large mesure arbitraires : il existe par exemple de très nombreuses manières de réaliser une opposition entre consonnes voisées et consonnes non-voisées, comme le montrent les différences bien connues sur ce point entre l'anglais et le français. Ces frontières

sont assujetties à des conventions partagées, tacitement établies par des générations entières de locuteurs. Lorsque nous accomplissons une opération aussi élémentaire que celle qui consiste à reconnaître une consonne voisée, les innombrables interactions inter-individuelles dont ces conventions sont le produit font écho en nous. Dès son premier niveau, la perception de la parole est intrinsèquement sociale.

Du traitement des sons de la parole à l'identité perçue du locuteur

Depuis quelque trente ans, les études se sont accumulées qui ont fait apparaître que la perception de la parole était influencée par différents facteurs de nature sociale. Elles ont montré que les auditeurs ne traitent pas les sons de la parole de la même façon selon la représentation que ces auditeurs se forment de l'identité sociale du locuteur. En manipulant cette représentation de manière expérimentale, les chercheurs ont constaté que le traitement de la parole est soumis à l'influence d'une large variété de caractéristiques attribuées au locuteur par l'auditeur, et notamment le genre, l'âge, la classe sociale. Il est également apparu que les propriétés dites indexicales dans le flux de parole, c'est-à-dire celles qui sont liées à l'identité individuelle et sociale du locuteur, ont un impact sur la reconnaissance des mots parlés. Ces découvertes ont profondément modifié notre conception des représentations mentales associées à la forme sonore du langage chez l'humain.

Dans les années 1990, de premières expériences ont révélé que les performances des auditeurs se dégradent lorsqu'ils avaient à reconnaître une série de mots produits par plusieurs locuteurs plutôt qu'un seul. On a également mis en évidence qu'il était plus difficile de reconnaître un mot que l'on a déjà entendu une première fois dans l'expérience lorsque le locuteur change entre la première et la seconde répétition. Ces résultats ont été interprétés en faveur de l'hypothèse selon laquelle les caractéristiques phonétiques propres au locuteur dans le flux de parole sont encodées par l'auditeur dans sa mémoire à long terme, et que cela requiert davantage de temps et de ressources cognitives lorsque le nombre de locuteurs entendus augmente. On a montré aussi qu'il était possible de faire bouger la frontière établie par l'auditeur entre deux phonèmes, tels que /s/ et /ʃ/, selon que l'on présentait ces sons comme ayant été produits par un homme ou une femme à l'auditeur. D'autres études ont fait apparaître que les voyelles étaient catégorisées différemment selon l'origine régionale et la classe sociale associées au locuteur pour l'auditeur. Des modèles nouveaux de la reconnaissance des mots parlés ont alors été développés, dans lesquels les propriétés indexicales font partie intégrante des représentations associées à la forme sonore des mots parlés dans le lexique mental.

Converger vers autrui dans la communication parlée

Dans les recherches que nous menons aujourd'hui sur la production et la compréhension du langage oral, s'observe une tendance majeure qui est de replacer le langage oral à l'intérieur de son site d'occurrence premier, les interactions sociales. Et dans le décours d'une interaction conversationnelle, il a été montré que chaque participant a tendance à imiter, ou encore à converger, vers l'autre participant. Les phénomènes de convergence ou d'imitation se manifestent sous des formes d'une grande variété, au travers de la posture, des mouvements de

la tête, des expressions faciales, et de la parole. On considère qu'ils ont pour effet de rendre l'échange conversationnel plus facile en contribuant à la mise en place d'un socle commun entre les interlocuteurs. Ils présentent un lien avec les mécanismes d'alignement dont les représentations linguistiques formées par les différents interlocuteurs sont supposées être le siège, au niveau phonologique, syntaxique et sémantique, dans le but de rendre le modèle de la situation propre à chaque interlocuteur mieux accessible aux autres.

Dans les recherches menées en psycholinguistique, ces effets de convergence ont été assimilés à une sonde utilisable pour explorer la plasticité des représentations associées à la forme sonore des mots dans le lexique mental, chez des auditeurs exposés à la voix d'une autre personne. Selon Goldinger (2000), « the imitation data verify that the contents of memory can be reflected in the sound of a person's voice ». À partir du milieu des années 2000, des expériences ont été réalisées dans lesquelles différents locuteurs ont été amenés à produire les mêmes mots avant, pendant et après une conversation. Les enregistrements ont ensuite été donnés à entendre à des auditeurs, qui ont jugé que la similarité présentée d'un locuteur à l'autre par les mots dans leur forme prononcée s'était accrue au fil de l'interaction, et persistait après que cette interaction eut pris fin. Ces travaux ont eu de considérables répercussions, car ils laissent à penser que les interactions conversationnelles ont un impact durable sur la forme prononcée des mots, et que les mots sont stockés en mémoire sous une forme qui est étroitement assujettie à leur contexte d'utilisation. Selon cette hypothèse, les représentations associées à la forme sonore des mots dans le lexique mental évoluent au fil d'une conversation sous l'influence de la manière de parler de l'interlocuteur, et conservent des traces de cette influence une fois que la conversation a pris fin. Une approche nouvelle du traitement de la parole s'est ainsi formée, dans laquelle les représentations mises en jeu dans le traitement de la parole revêtent un caractère dynamique, et se transforment au fil de la vie de chacun sous l'influence de la parole produite par autrui.

Pour aller plus loin

Alario, F.-X., Dufour, S., & Pattamadilok, C. (2014). Psychologie cognitive du langage. *Lettre des Neurosciences*, 46 : 9–10.

Moulin-Frier, C., Diard, J., Schwartz, J.-L., & Bessière, P. (2015). COSMO ("Communicating about Objects using Sensory-Motor Operations"): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems. *Journal of Phonetics*, 53:5–41.

Nguyen, N. (2012). Representations of speech sound patterns in the speaker's brain: Insights from perception studies, in Cohn, A., Fougeron, C., & Huffman, M., éditeurs, *Handbook of Laboratory Phonology* (Oxford University Press, Oxford, UK), pp. 359–368.

Tuller, B., & Nguyen, N. (2012). Acoustic and phonological learning, in Seel, N.M., éd., *Encyclopedia of the Sciences of Learning* (Springer Verlag), pp. 51–54.