

Psycholinguistique - Contrôle Continu

Brice Boucard

Master 1, Semestre 2 - Mai 2006

Compte rendu de :

Noël Nguyen, « Perception de la parole »

in N. Nguyen, S. Wauquier-Gravelines et J. Durand

Phonologie et Phonétique : Forme et Substance,

chap. 16, pp. 425-447, Paris : Lavoisier.

Noël Nguyen commence par rappeler le paradoxe qui se trouve en phonologie et en phonétique ; en effet, ces deux “disciplines” sont surtout abordées, fondées sur la production et notamment « *sur des critères de nature articulatoire* », alors même que « *la forme sonore du langage est d’abord appréhendée à travers l’oreille* ». D’où la nécessité d’une bonne connaissance de ce système perceptif qu’est l’audition puisque son fonctionnement, « *[s]es propriétés [...] conditionnent la façon dont nous appréhendons les phénomènes phonétiques et phonologiques* ». Au-delà de cet enjeu, N. Nguyen affirme que de telles recherches permettent également de se poser également la question de la place de l’auditeur dans les théories que proposent la phonétique ou la phonologie et d’interroger les « *relations qui s’établissent entre phonétique, phonologie et cognition* ». Il fait alors un bref rappel des théories qui se basent sur ou qui font référence à la perception et cite les théories proposées par Jakobson, Chomsky et Halle, ainsi que les théories plus récentes que sont la théorie de l’optimalité et la théorie AIU.

N. Nguyen se propose dans ce chapitre « *de présenter un aperçu général des recherches sur la perception de la parole, dans leur relation avec la phonétique et la phonologie* » et va pour ce faire commencer par présenter comment se fait l’identification des phonèmes avant d’aborder « *les questions relatives à la forme et à la fonction des représentations phonétiques et phonologiques dans le traitement de la parole* ».

Comment l'auditeur identifie-t-il les phonèmes ?

Les recherches ont pendant longtemps porté sur les frontières entre phonèmes : dès les années cinquante, des études ont été réalisées pour identifier les indices responsables des oppositions entre phonèmes ; ces études ont permis de mettre au jour, notamment pour ce qui est de la perception des occlusives, la "perception catégorielle", phénomène qui, selon N. Nguyen, « renvoie au fait qu'il est, dans certaines circonstances, plus facile à l'auditeur de percevoir les différences entre deux sons lorsque ces derniers se rangent dans deux catégories phonémiques différentes, plutôt que dans une même catégorie ». La perception des phonèmes serait donc rendue possible du fait d'une certaine "indifférence" aux variations acoustiques intercatégorielles et de l'attention portée aux différences intercatégorielles. Des études ont été menées par la suite sur la perception des voyelles ; dans ce cas là, l'auditeur perçoit les différences entre voyelles, qu'elles soient ou non de la même catégorie phonémique. Il en a, donc, été déduit que le traitement des consonnes et des voyelles faisait appel à des processus, à « deux modes de traitement différents ». La théorie motrice de la perception de la parole, « dans laquelle un lien étroit est postulé entre perception et production de la parole » et plus particulièrement entre perception et mouvements articulatoires, a trouvé dans cette dichotomie un argument majeur puisque, selon ces chercheurs, « il n'y a pas de frontières bien établies entre voyelles dans l'espace articulatoire » ce qui nous permet de produire deux voyelles différentes l'une après l'autre sans rupture, « de manière continue et graduelle » alors que « les consonnes occlusives seraient séparées par par un ensemble de frontières articulatoires naturelles », c'est-à-dire qu'il est impossible de produire deux consonnes différentes l'une après l'autre sans rupture. Ainsi, notre conduit vocal serait à la base d'une telle différence dans la perception des consonnes et des voyelles, les premières étant perçues de manière catégorielle, les secondes de façon continue. Cependant, Fujisaki et Kawashima ont montré que la perception des fricatives,glides et liquides n'était pas aussi catégorielles que celle des occlusives ainsi que les voyelles pouvaient, dans certaines conditions, être perçues de manière catégorielle ; selon eux, la possibilité d'adopter, en perception de la parole, un processus catégoriel ou non serait due à la mémoire auditive à court terme. Selon Massaro, si l'auditeur, lors d'expérimentations, peut donner une réponse catégorielle, cela ne signifie pas qu'il perçoit de façon catégorielle ; en effet, l'auditeur percevrait la parole de manière continue, « conserv[ant ainsi] autant d'information que possible sur le signal le plus longtemps possible », ce qui permet de corriger d'éventuelles erreurs de compréhension, d'identification de phonèmes ou de mots.

Noël Nguyen conclut en proposant que « l'auditeur puisse procéder à un

double codage de l'information, continu (sous la forme d'une sorte de spectrogramme auditif par exemple) et discret (sous la forme d'un ensemble d'étiquettes phonétiques) »- cette dernière forme de codage étant à mettre en relation avec la perception catégorielle - dont « l'importance relative [...] pourrait dans une certaine mesure dépendre de la tâche assignée au sujet ».

Après avoir longtemps considéré que les différences intracatégorielles n'étaient pas perçues par les auditeurs, les linguistes se sont finalement penché « sur l'organisation interne de chaque catégorie », à la suite de Joanne Miller, puisque « selon [lui] les catégories phonémiques sont dotées d'une véritable structure interne, riche et graduée ». Des expériences ont ainsi montré que certains stimuli sont plus représentatifs de la catégorie phonémique à laquelle ils appartiennent, donc qu'il est possible de hiérarchiser les stimuli correspondant à une même catégorie ; le stimulus qui est considéré comme le meilleur exemplaire de la catégorie en constitue le prototype. Si on a d'abord considéré que la décision d'appartenance d'un son de parole à telle ou telle catégorie - unique - se « bas[ait] sur une série de décisions binaires », on considère aujourd'hui qu'une telle identification se base sur un processus probabiliste.

Kuhl, dans sa théorie des aimants perceptifs, fait du prototype l'élément central de la catégorisation des sons de parole puisque « selon [elle], les “bons” exemplaires, ou prototypes, ont un rôle central à jouer dans la structuration perceptive de l'espace phonétique ». Pour elle, les prototypes constituent des noyaux qui attireraient les sons qui lui ressemblent (pour ce qui est de leurs caractéristiques physiques), rendant ainsi difficile pour l'auditeur la distinction des différences qui existent entre ces sons. De là découle une nouvelle “conception” des frontières catégorielles qui se trouvent être fondées sur les sons prototypes dans la mesure où « La perception d'une frontière entre catégories est alors attribuée au simple fait que les sons situés à mi-chemin entre deux prototypes échappent à l'attraction de ces prototypes ».

Le contexte (contexte phonétique et contexte lexical) a-t-il une influence sur l'identification phonémique ? Le contexte phonétique renvoie essentiellement à la coarticulation. De nombreux auteurs considèrent que ces phénomènes de coarticulation ne gênent pas l'auditeur dans l'identification phonémique ; au contraire même : l'auditeur se servirait de ce contexte, des variations de réalisation pour faciliter l'identification « dans la mesure où les effets de contexte revêtent un caractère systématique et régulier ». Certains auteurs affirment que « l'auditeur ajuste les frontières qu'il établit entre phonèmes en fonction du contexte » : c'est la compensation perceptive. Qu'en est-il du contexte lexical ? Des expériences

ont montré que l'auditeur peut utiliser ses connaissances lexicales pour « *reconstituer la forme sonore d'un phonème remplacé par du bruit dans une phrase* » : il s'agit de la restauration phonémique. Le lexique intervient donc dans des processus de niveau inférieur : « *on parle d'un effet descendant du lexique sur la perception de la parole* » ; cependant, de tels effets ne font pas l'unanimité.

Comment l'auditeur se représente-t-il le signal de parole à chaque étape de traitement ?

Noël Nguyen commence par rappeler que, bien souvent, on considère que « *le signal de parole est décomposé sous la forme d'une séquence d'unités perceptives de base (ex. : phonèmes) à partir desquelles il est possible à l'auditeur d'accéder à des unités linguistiques à la fois plus larges (ex. : mots) et plus petites (ex. : traits distinctifs)* » ; ces unités perceptives se situeraient donc à un niveau infralexical. Plusieurs études ont montré que la syllabe constituait une unité perceptive de base ; cependant, des études postérieures à ces dernières tendent à prouver que cet “effet syllabique” est peut-être propre à certaine langue (dont le français), et que, par conséquent, la structure phonologique des langues “contraint” la perception de la parole. Si l'effet syllabique n'a pas été reproduit dans des expériences récentes, celles-ci semblent montrer que certains constituants syllabiques ont un rôle à jouer dans la perception de la parole et dans l'accès au lexique. D'autres chercheurs soutiennent que l'unité perceptive de base serait le trait, et que « *l'auditeur établit une correspondance directe entre une représentation en traits du signal de parole et le lexique, sans passer par un niveau phonémique ou syllabique* ». Certains résultats « *vont dans le sens d'un accès lexical direct à partir du trait* ». Quelle est donc l'unité perceptive de base ? Certains auteurs ont proposé que « *des unités de petite et de grande taille soient simultanément activées, avec un biais naturel en faveur des unités les plus larges* » et que ce “choix” dépende de divers facteurs.

Le signal de parole est soumis à d'importantes variations ; comment l'auditeur y fait-il face ? A cette question tentent de répondre différentes hypothèses contradictoires que N. Nguyen regroupe schématiquement en deux types de modèles :

- les modèles abstractionnistes pour lesquels « *l'auditeur associe à chaque mot une représentation phonologique abstraite et indépendante des caractéristiques individuelles du locuteur* ». L'un des arguments en faveur de cette approche abstractionniste est le fait que « *les mécanismes de traitement de la parole semblent très résistants au bruit et à la variabilité intra-*

et inter-individuelle ». Le modèle FUL - *Featurally Underspecified Lexicon* - est le plus radical de cette approche : « *dans ce modèle, chaque morphème est associé à une représentation phonologique, unique, abstraite et sous-spécifiée, indépendamment des variations dont ce morphème peut être le siège en surface* ». Selon certains chercheurs, les variations en surface ne seraient pas perçues par l'auditeur, notamment pour ce qui est des variations dues au phénomène d'assimilation, ce qui serait permis du fait que certains traits « *ne s[eraient] pas spécifiés dans le lexique* » ; ainsi, grâce à cela, l'assimilation ne serait pas détectée par l'auditeur. Dans ces modèles abstractionnistes, un processus de normalisation précoce dans le traitement du signal intervient afin de réduire les variations inter-individuelles. Ces modèles reposent sur une « *perspective dualiste postulant l'existence de deux niveaux de représentation : une représentation de surface, physique et variable, et une représentation phonologique sous-jacente, abstraite et symbolique* », ces représentations phonologiques sous-jacentes étant « *pré-déterminées génétiquement* » et stimulées et stabilisées par l'input que les enfants reçoivent. Cette approche propose donc un traitement de la parole qui se ferait par réduction de la quantité d'information retenue par l'auditeur, basé sur le postulat d'une mémoire de taille limitée - ce qui a été "démenti" par la suite.

- C'est justement la grande capacité de la mémoire humaine qui a permis de penser aux modèles à exemplaires, dans lesquels « *l'auditeur se représente mots et constructions grammaticales de manière concrète et détaillée, sous la forme de listes d'exemplaires et/ou de prototypes* », listes conservées en mémoire puisque ces modèles « *attribuent à l'auditeur la capacité de stocker en mémoire les différentes formes de surface associées à un mot* ». Selon ces modèles, la « *variabilité est encodée par l'auditeur au niveau lexical* », toutes les formes de surface correspondant à un mot étant stockées en mémoire et l'identification d'un mot se faisant par comparaison, analogie entre la forme entendue et tous les exemplaires stockés. L'auditeur identifiera le signal de parole en repérant le mot pour lequel le plus grand nombre d'exemplaires aura été activé. Ces modèles sont de plus dynamiques - dans la mesure où les "prototypes" créés par l'activation des exemplaires - ne cessent d'être mis à jour et ce, en temps réel ; de plus, « *se montrent [...] capables de segmenter le signal de parole sous la forme d'une séquence d'unités infra-lexicales (traits, phonèmes, syllabes)* ». Les modèles à exemplaires se trouvent fondés par certains résultats expérimentaux, tels que le fait que « *l'auditeur est sensible à la structure détaillée du signal acoustique dans la compréhension de la parole* » et que des paramètres phonétiques fins entrent en jeu dans l'identification des mots ;

« *l'information extraite par l'auditeur à partir du signal de parole dans l'accès au lexique présent[erait donc] un caractère riche et gradué* ». L'auditeur serait également « *sensible aux caractéristiques individuelles de la voix du locuteur* », un mot étant plus facilement identifié si l'auditeur l'a déjà entendu de la bouche du même locuteur. Cependant, ces modèles à exemplaires ne font pas l'unanimité dans la mesure où ils révèlent certaines contradictions et délaissent certaines questions.