

Un modèle psycholinguistique informatique de la reconnaissance des mots dans la chaîne parlée du français

NICOLAS LÉWY, FRANÇOIS GROSJEAN,
LYSIANE GROSJEAN, ISABELLE RACINE ET
CAROLE YERSIN

Laboratoire de traitement du langage et de la parole, Université de Neuchâtel

(Received March 2004; accepted October 2004)

ABSTRACT

Nous proposons un nouveau modèle psycholinguistique informatique du français. Le modèle, intitulé FN5, porte sur la reconnaissance de mots parlés, présentés en isolé (déterminant, adjectif antéposé, substantif) ou en suites de deux mots (déterminant et substantif, adjectif antéposé et substantif). Grâce à un lexique de plus de 17 000 mots, à une architecture connexionniste localiste, et à des mécanismes développés spécifiquement (processeur de position, groupements de connexions, et point d'isolation), le modèle peut simuler, de manière séquentielle ou simultanée, certains effets propres au mot isolé (fréquence, longueur, homophonie) ou à une suite de mots (coupure lexicale, enchaînement, liaison, effacement du schwa). De plus, le modèle tient compte de certaines différences qui existent entre le français standard et le français de Suisse romande (nombre de voyelles, durée vocalique en fin de mot, et statut du schwa). Nous décrivons ce modèle, présentons son interface graphique, et l'illustrons avec des exemples de simulation.

INTRODUCTION

Une des questions les plus centrales de la psycholinguistique de l'adulte concerne la reconnaissance de la forme des mots dans la chaîne parlée. Elle a lieu environ 175 fois par minute lors d'une conversation à débit normal et elle nous permet d'avoir accès à l'ensemble des informations d'une entrée lexicale qui sont nécessaires au traitement de la parole (informations morphologiques, syntaxiques, sémantiques, etc.). C'est en grande partie grâce à notre lexique mental que les niveaux supérieurs d'analyse peuvent participer à la construction de la représentation interprétative (le sens enrichi) de l'énoncé. Les études expérimentales qui traitent de la reconnaissance de mots parlés existent depuis au moins une centaine d'années (voir, par exemple, Bagley, 1900, pour une première étude). Il s'avère que la modélisation de la reconnaissance de mots parlés est plus récente, et qu'au départ, les psycholinguistes

ont proposé des modèles dont la description était verbale (ex. Morton, 1969; Marslen-Wilson et Welsh, 1978; Grosjean et Gee, 1987; Marslen-Wilson, 1987, etc.).

Depuis une quinzaine d'années, les modèles informatiques ont fait une entrée remarquable en psycholinguistique. Cherchant à simuler sur ordinateur le traitement du langage chez l'être humain (sa perception, compréhension, production et/ou mémorisation), ils appartiennent à une nouvelle branche de recherche, la psycholinguistique informatique (voir des ouvrages généraux tels que Crocker, 1996; Dijkstra et De Smedt, 1996; Plunkett et Elman, 1997, etc.). Les avantages d'une modélisation informatique sont nombreux. Alors que la modélisation verbale est forcément imprécise, incomplète, et souvent ambiguë, la modélisation du traitement humain à l'aide de l'informatique force le chercheur à rendre explicite tous les mécanismes du modèle. Outre la possibilité de visualiser le processus psycholinguistique en question, ce type de modèle permet un va-et-vient fructueux entre données expérimentales et théorisation. En conséquence, la modélisation informatique est devenue une partie intégrale, et indispensable, de la psycholinguistique.

Parmi les divers modèles informatiques qui ont été proposés pour simuler la reconnaissance de mots parlés (le processus psycholinguistique qui nous intéresse ici), c'est le modèle TRACE de McClelland et Elman (1986) qui a eu le plus d'impact dans la littérature à la fois théorique et expérimentale. De plus, TRACE a fait l'objet de nombreuses études de simulations et d'évaluations (ex. Frauenfelder et Peeters, 1998). Signalons aussi le modèle SHORTLIST (Norris, 1994), celui de Gaskell, Hare et Marslen-Wilson (1995), et les modèles MERGE (Norris, McQueen et Cutler, 2000) et PARSYN (Luce, Goldinger, Auer et Vitevitch, 2000).

Malgré l'apport incontestable des modèles existants, il se trouve que tous ceux ayant trait à la reconnaissance de mots parlés traitent de l'anglais.¹ Or, il y a un réel besoin de modèles qui tiennent compte de la spécificité des différentes langues. Par exemple, pour le français, outre la simulation d'effets propres au mot en isolé que l'on retrouve dans plusieurs langues (ex. la fréquence d'occurrence du mot, sa longueur, l'homophonie qui existe entre certains mots, etc.), le modèle doit pouvoir rendre compte notamment de la reconnaissance de suites de deux mots liés par un enchaînement avec ou sans liaison (ex. 'petit ami' et 'chaque ours') ou contenant un effacement du schwa (ex. 'la s'maine'). Il serait heureux également qu'il tienne compte de différentes prononciations régionales (ex. la présence de voyelles finales longues en français de Suisse romande, comme dans 'amie').

Dans cet article, qui comprend trois parties, nous présentons une première version d'un modèle psycholinguistique informatique du français, modèle appelé FN5. Dans la première partie, qui décrit le modèle, nous familiarisons le lecteur avec l'interface et offrons un premier exemple de son utilisation. Nous présentons son lexique et nous traitons de son architecture en mettant en avant trois caractéristiques

¹ Ans, Carbonnel et Valdois (1998) ont certes proposé un modèle informatique pour le français, mais celui-ci concerne la lecture de mots écrits.

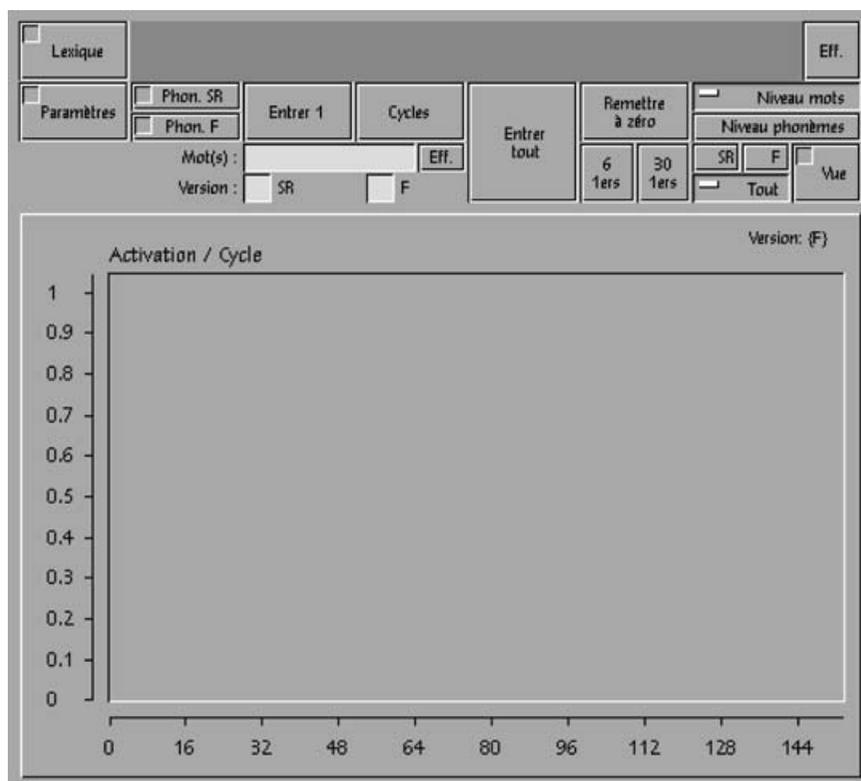


Figure 1. Fenêtre principale de l'interface (niveau des mots).

qui lui sont propres: le processeur de position, les groupements de connexions, et le point d'isolation. Dans la deuxième partie, nous présentons divers exemples de simulation, et dans la troisième partie, nous donnons des détails sur quelques autres aspects du modèle.

I LE MODÈLE

1.1 Interface et illustration

En démarrant le logiciel, l'utilisateur du modèle se trouve devant une interface avec fenêtres et éléments graphiques. La fenêtre principale est montrée dans la Figure 1. Comme on le voit, elle est composée de deux parties: celle du haut, avec divers boutons, zones de texte, etc., sert à régler les différentes composantes du modèle et à entrer des mots pour lesquels on simule la reconnaissance; celle du bas, libellée 'Activation/Cycle', sert à visualiser l'activation des mots.

Afin d'illustrer le modèle en activité, nous allons suivre la reconnaissance du premier mot de la suite 'petit ami' (/pitami/) prononcée avec un effacement du

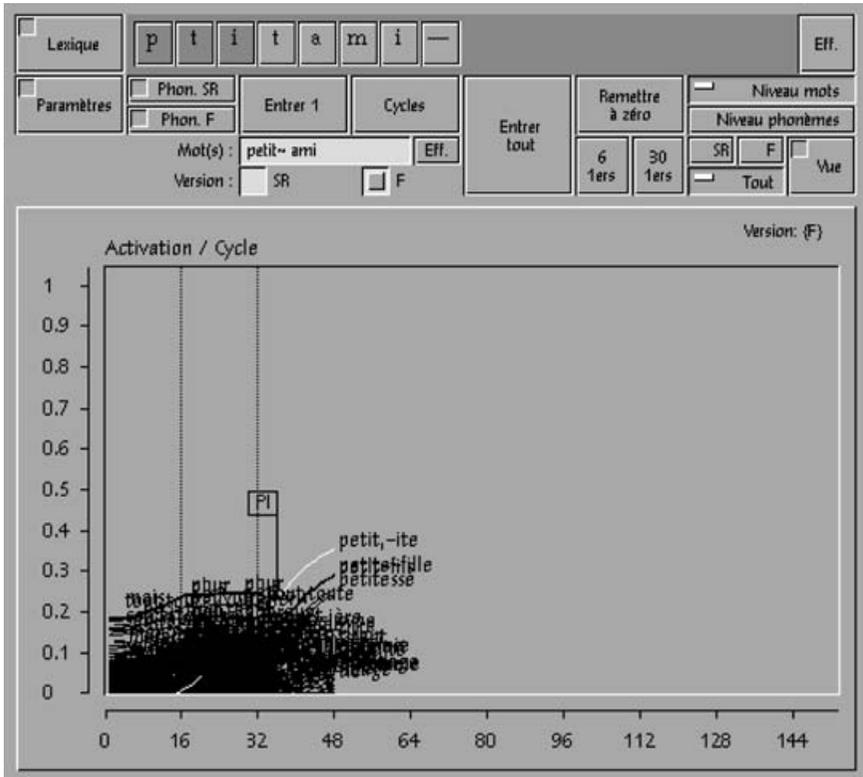


Figure 2. Fenêtre de l'interface (niveau des mots) avec en entrée 'petit ami'; le taux d'activation des mots est celui obtenu après les trois premiers phonèmes (/p/, /t/ et /i/).

schwa dans 'petit' et une liaison entre 'petit' et 'ami'. Nous devons d'abord charger le lexique complet du modèle dans sa version française en cliquant sur la petite case du bouton 'Lexique' (en haut à gauche). Une fenêtre s'ouvre et nous permet ainsi de choisir un fichier lexical. Ensuite, nous tapons les mots 'petit ami' dans la zone de texte prévue à cet effet (intitulée 'Mot(s)'). La transcription phonétique /ptitami-/ est alors fournie par le logiciel dans la zone en haut, à droite du bouton 'Lexique', comme on le voit dans la Figure 2. Notons que dans la suite phonétique qui apparaît, le /-/ marque la fin de la chaîne. Quant à la liaison, elle se manifeste visuellement de deux manières différentes: au niveau de la suite phonétique, un /t/ est inséré automatiquement au bon endroit, et dans la suite graphique (zone 'Mot(s)'), un signe '~' apparaît entre 'petit' et 'ami' lorsque l'on fait un retour à la ligne après avoir tapé les mots.

Pour traiter les trois phonèmes du premier mot de la suite (à savoir, /p/, /t/ et /i/), nous cliquons trois fois sur le bouton 'Entrer 1', bouton qui sert à insérer les phonèmes dans le modèle, un par un, en commençant avec le premier phonème. L'état d'activation des mots après l'entrée des phonèmes apparaît dans la zone

'Activation/Cycle' de la Figure 2. Certes, au tout début du processus (à gauche donc de la fenêtre), il y a de nombreux candidats possibles, d'où la masse de mots actifs, mais on constate qu'à la fin des trois phonèmes (au niveau du cycle 48 donc), le mot 'petit, -ite' est déjà le plus actif de tous les candidats possibles (voir la courbe blanche qui représente le niveau d'activation du mot entré, le 'mot cible'). A tout moment, lors de la simulation, il est possible d'obtenir une liste des candidats les plus actifs, et leur taux d'activation, en cliquant sur le bouton '6 1ers' (six premiers) ou '30 1ers' (trente premiers).

Une fois que l'utilisateur a étudié l'information qui lui est fournie à chaque étape (dans notre cas, après trois phonèmes), il peut poursuivre en cliquant sur 'Entrer 1'. Dans notre cas, il faudrait cliquer encore cinq fois afin de présenter au modèle les phonèmes restants y compris la pause de fin de suite (/tami-/). Une autre approche que l'on peut utiliser au tout début de la simulation est de cliquer sur 'Entrer tout', ce qui fait que la chaîne complète est insérée, phonème par phonème, mais sans que la simulation ne s'arrête après chaque phonème. (Nous reviendrons sur la suite 'petit ami' plus loin dans l'article, et verrons l'état d'activation à sa fin).

A tout moment, on peut cliquer sur le bouton en haut à droite, 'Niveau phonèmes', afin de visualiser l'état d'activation des phonèmes. La Figure 3 présente celui des trois premiers phonèmes de 'petit ami': les carrés noirs représentent l'activation et les carrés blancs l'inhibition alors que le niveau d'activation ou d'inhibition est indiqué, lui, par la taille du carré; la légende des phonèmes est à gauche. Nous observons que le modèle a activé les phonèmes auxquels on s'attendait et dans le bon ordre. Ici aussi, on peut cliquer sur '6 1ers' ou '30 1ers' pour obtenir la valeur d'activation des phonèmes les plus actifs.

A l'aide de la fenêtre principale du logiciel, et de plusieurs fenêtres secondaires, l'utilisateur accède à d'autres fonctionnalités du modèle (ex. réglage des paramètres, gestion du lexique, option de cycles supplémentaires, remise à zéro de toutes les unités, réglage de la visualisation, options pour entrer des mots avec et sans liaison, avec et sans pause, etc.).

1.2 Bases de données linguistiques

Le modèle contient un lexique substantiel de 16 971 substantifs, et 697 déterminants et adjectifs antéposés, pour un total de 17 668 mots, et ce dans deux versions, française et Suisse romande. Pour la version Suisse romande, nous avons ajouté aux 35 phonèmes du français standard le /a/, le /œ/, et sept voyelles longues, /i: e: y: ø: u: α:/, afin d'obtenir 44 phonèmes en tout (Métral, 1977). Pour établir le lexique, nous avons tout d'abord extrait l'ensemble des déterminants, adjectifs et substantifs de la base de données BRULEX (Content, Mousty et Radeau, 1990) afin de les soumettre à une série de traitements à la fois informatiques et linguistiques. Nous avons éliminé les mots de certaines catégories (ex. substantifs au pluriel, adjectifs postposés uniquement, variantes orthographiques, doublons, mots de très basse fréquence, etc.), nous avons normalisé les indices de fréquence des mots (les valeurs s'étendent désormais de 0 pour très rare à 1 pour très fréquent), et nous avons vérifié

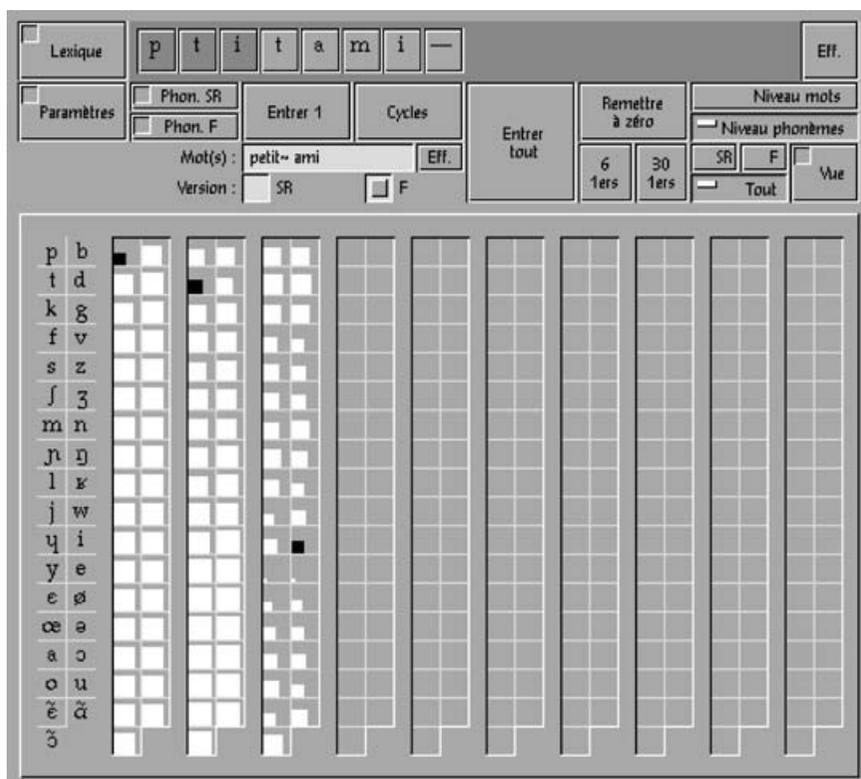


Figure 3. Fenêtre de l'interface (niveau des phonèmes) avec en entrée 'petit ami'; le taux d'activation des trois premiers phonèmes de la suite correspond aux carrés noirs.

toute l'information linguistique présente dans le lexique (orthographe, transcription phonétique et genre). Nous avons également ajouté des informations sur les voyelles longues en finale de mot pour le français de Suisse romande (ex. 'amie', 'vraie', etc.) et sur les consonnes de liaison des adjectifs antéposés (ex. 'grand' avec /t/ obligatoire, 'gras' avec /z/ facultatif, etc.). Enfin, nous avons présenté les mots qui comportent un schwa, qu'il soit obligatoire dans la prononciation, facultatif ou interdit (2 190 mots en tout), à 12 locuteurs du français de Suisse romande et 12 locuteurs du français standard (région de Nantes), afin d'obtenir un indice de préférence de la prononciation avec et sans effacement du schwa (ex. 'f'nêtre' et 'fenêtre')², et nous avons ajouté les moyennes obtenues aux deux versions du lexique.

² Voir la thèse en préparation d'Isabelle Racine, "Les effets de l'effacement du schwa sur la perception et la production de la parole en français".

1.3 Architecture et fonctionnement

Le modèle se sert du formalisme des réseaux connexionnistes localistes, approche proposée par McClelland et Rumelhart (1981) et qui a reçu une attention renouvelée ces dernières années (pour une discussion des mérites de l'approche, voir, par exemple, Grainger et Jacobs, 1998; Page, 2000).

Dans la Figure 4, nous présentons l'architecture générale du modèle. Celle-ci s'inspire en partie des modèles informatiques actuels les plus influents de la reconnaissance de mots parlés, notamment TRACE (McClelland et Elman, 1986) et SHORTLIST (Norris, 1994). Le modèle consiste en un grand nombre d'unités de base qui travaillent toutes en parallèle. Elles additionnent des signaux en entrée, signaux positifs ('d'activation') et/ou négatifs ('d'inhibition'), et elles envoient, selon certaines conditions, des signaux similaires en sortie à d'autres unités du modèle. Comme nous le voyons dans la figure, ces unités sont organisées en trois niveaux linguistiques – traits distinctifs, phonèmes et mots –, et sont connectées par des liens servant à transmettre les signaux, ces liens étant des connexions d'activation (visualisées par les flèches pointues) et/ou des connexions d'inhibition (visualisées par les flèches rondes). Il existe des liens ascendants, des traits distinctifs vers les phonèmes ainsi que des phonèmes vers les mots, et des liens descendants, des mots vers les phonèmes (optionnels dans le modèle; voir ci-dessous). De plus, il y a des liens latéraux, aux niveaux des phonèmes et des mots.³

A titre d'exemple, que se passe-t-il lorsque l'on présente au modèle, en entrée, la chaîne phonétique /tas/ qui correspond au mot 'tasse'? Les traits distinctifs, tout d'abord, envoient des signaux d'activation et d'inhibition vers les phonèmes. Cela active /t/, /a/ et /s/, dans les positions 1, 2 et 3 respectivement; cela active un peu moins les phonèmes qui leur sont phonétiquement proches (ex. /d/ pour /t/), et cela inhibe les phonèmes plus distants (ex. /f/, /ø/, etc.). Dès que les phonèmes sont suffisamment activés, ou inhibés selon le cas, ils envoient à leur tour des signaux d'activation, ou d'inhibition, vers les mots. Donc, /t/, /a/ et /s/ activent 'tasse' ainsi que d'autres mots qui contiennent certains de ces phonèmes dans les mêmes positions séquentielles (ex. 'ta', 'tard', 'masse', etc.). Etant donné que le mot 'tasse' contient les trois phonèmes concernés, dans l'ordre approprié, il reçoit l'activation la plus élevée. Les phonèmes absents de la chaîne en entrée (ex. /f/ et /ø/) inhibent les mots qui contiennent ces mêmes phonèmes (ex. 'feu'). Dès que les mots sont suffisamment activés, ou inhibés, ils renvoient à leur tour des signaux d'activation, ou d'inhibition, vers les phonèmes, si les mécanismes respectifs sont enclenchés (voir plus haut). Dans ce cas, 'tasse', parmi d'autres, active, et ainsi renforce, /t/, /a/ et /s/, dans les positions concernées. De même, les mots absents (ils sont bien entendu fort nombreux) affaiblissent les phonèmes dont ils sont composés. En parallèle, chaque phonème activé inhibe les autres phonèmes (ex. /t/ inhibe /d/), et chaque mot activé inhibe les autres mots (ex. 'tasse' inhibe 'tard'), réduisant

³ Afin de faciliter la présentation, nous utilisons les raccourcis 'traits distinctifs', 'phonèmes' et 'mots', pour parler des unités connexionnistes qui servent dans le modèle à représenter les unités linguistiques équivalentes.

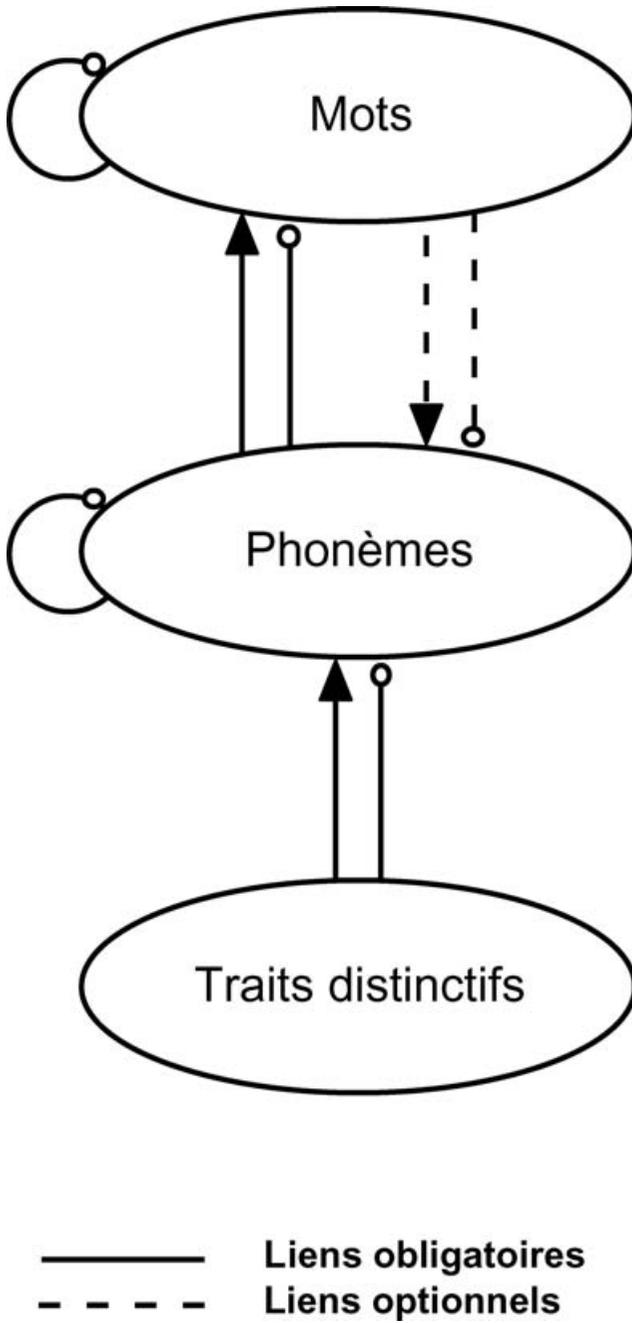


Figure 4. Architecture générale du modèle.

ainsi systématiquement le nombre d'unités encore influentes. A la fin du processus, l'état du modèle est tel que l'on trouve activées le plus fortement les seules unités appropriées, que cela soit au niveau des phonèmes (/t/, /a/ et /s/) ou au niveau des mots ('tasse').

Dans ce qui suit, nous décrivons les trois caractéristiques développées spécifiquement pour le modèle.

1) *Processeur de position*

Afin de permettre la simulation de la reconnaissance d'une suite de deux mots (ex. 'ta tasse' - /tatas/), nous avons prévu qu'à chaque cycle, un processeur de position aligne les mots du lexique avec chaque position possible de la chaîne des phonèmes en entrée. Dans le cas du mot 'table', par exemple, le modèle aligne le début du mot avec le début de la suite, à savoir, dans l'exemple /tatas/, avec le /t/ initial en position 1, mais également avec le /a/ en position 2, avec le deuxième /t/ en position 3, etc. Il établit pour chacune des positions la somme combinée des activations et inhibitions du mot (signaux ascendants ainsi que latéraux), compare ces sommes, et choisit la position qui produit la somme la plus élevée pour ce mot. C'est cette position qui est alors retenue et stockée avec le mot pour la suite de la simulation. Ce positionnement optimal du mot est répété à chaque cycle d'un phonème (il y a 16 cycles par phonème) et cela pour chaque phonème de la chaîne phonétique. Certes, la position du mot peut changer d'un cycle à l'autre, mais nous imposons au système de n'en avoir qu'une seule, pour chaque mot, à chaque cycle. Si plusieurs positions donnent exactement le même résultat, le processeur de position choisit la position la plus récente. Bien entendu, dans un traitement séquentiel où les phonèmes arrivent les uns après les autres (d'abord le premier phonème, ensuite le deuxième phonème, etc.), le processeur de position ne peut placer un mot que dans la position 1 lors de l'arrivée du premier phonème, et aucune comparaison n'est effectuée à ce moment-là. C'est à l'arrivée du deuxième phonème qu'il peut placer le mot en position 1 ou en position 2 en comparant les deux sommes, et c'est à l'arrivée du troisième phonème qu'il compare les sommes des positions 1, 2 et 3, etc. De la sorte, l'activation et la reconnaissance d'un mot peuvent avoir lieu à n'importe quel emplacement de la chaîne parlée.

Continuons l'illustration du processus avec la chaîne phonétique en entrée /tatas/ ('ta tasse') et avec le mot-candidat 'tasse' qui se trouve dans le lexique. Pendant l'arrivée du premier /t/, qui active des mots qui ont un /t/ initial (c'est le cas de /tas/), le mot 'tasse' occupe d'abord la position 1. En effet, il n'y a pas d'autres positions possibles à ce stade et le processeur de position n'a pas d'autres alternatives. Pendant l'arrivée du /a/, le processeur compare les sommes des activations et inhibitions de 'tasse' quand ce mot est aligné en position 1 (avec le /t/) et en position 2 (avec le /a/), et relève que la première des deux sommes est la plus élevée (en effet, /tas/ ne commence pas avec un /a/); il garde donc l'indication de la position 1 stockée avec le mot 'tasse'. Lorsque le deuxième /t/ arrive, le processeur de position compare les sommes pour les positions 1, 2 et 3, et constate, probablement après quelques cycles de ce deuxième /t/, que la somme de 'tasse'

en position 1 est devenue moins importante que celle en position 3. Ceci est dû à l'inhibition de 'tasse' en position 1 causée par le /s/ qui est absent, et donc inhibé, lors de l'arrivée du deuxième /t/. Le processeur de position stocke donc la position 3 avec 'tasse' à partir de cet instant. En fait, à ce moment-là, le mot 'tasse' a cessé d'être un candidat aligné au début de la chaîne phonétique (/ta---/) pour devenir un candidat aligné en position 3 (/--tas/). C'est justement ce qu'il faut car 'tasse' ne peut plus être un candidat pour /tat/ mais peut être un nouveau candidat pour la séquence qui commence avec le deuxième /t/. Pour la suite de la simulation, la position de 'tasse' demeure inchangée (position 3), car la comparaison avec les autres positions la donne toujours comme étant la meilleure. Entre-temps, le premier mot de la suite, 'ta', a accumulé de l'activation à son tour, surtout lorsque 'tasse' a été déplacé vers la position 3. A la fin de la simulation, 'ta', en position 1, et 'tasse', en position 3, se dégagent comme étant les meilleurs candidats pour la chaîne /tatas/. Le modèle vient donc de reconnaître une suite de deux mots.

Lorsque nous comparons cette approche avec celle d'autres modèles qui simulent la reconnaissance de deux ou plusieurs mots (ex. McClelland et Elman, 1986; Norris, 1994; Christiansen, Allen et Seidenberg, 1998), nous y voyons plusieurs avantages réunis en un seul modèle. Premièrement, l'approche avec processeur de position fonctionne sans qu'il faille dupliquer les mots à toutes les positions théoriquement possibles: les mots n'ont qu'une seule position qui leur est attribuée à chaque cycle et ils peuvent être déplacés vers une autre position si le processeur de position le décide. Deuxièmement, l'approche ne limite pas le nombre des mots-candidats qui sont pris en considération; tous les mots du lexique peuvent être activés en parallèle, et ce avec un niveau d'activation approprié (forte lors d'une bonne correspondance, moindre lors d'une correspondance partielle, etc.). Troisièmement, l'approche ne nécessite pas que la chaîne en entrée soit segmentée préalablement; elle s'appuie plutôt sur le fait que les frontières entre mots émergent souvent du processus normal de la reconnaissance des mots.

2) Groupements de connexions

Pour les mots qui contiennent des phonèmes qui peuvent être effacés (ex. un schwa) et/ou ajoutés (ex. une consonne de liaison), et plus généralement pour tous les mots avec plusieurs formes (comme les adjectifs), nous faisons appel aux groupements de connexions. Chacune des prononciations réalisables d'un mot (ex. avec ou sans schwa, avec ou sans consonne de liaison, masculin ou féminin, etc.) possède un groupement de connexions qui relie le mot avec les phonèmes en question. Par exemple, pour le substantif 'pelouse', il y a un groupement de connexions pour la forme avec effacement du schwa (/pluz/) et un autre pour la forme sans effacement (/pøluz/). De même, pour l'adjectif 'grand', il existe un groupement pour la forme féminine (/gʁãd/), un deuxième pour la forme masculine sans liaison (/gʁã/), et un troisième pour la forme masculine avec liaison (/gʁãt/).

Le processeur de position n'aligne donc pas seulement chaque mot avec chaque position possible dans la chaîne des phonèmes en entrée, mais aussi, dans le cas de plusieurs groupements de connexions, il établit les sommes des activations

et inhibitions pour chaque groupement, à chaque position, afin de trouver le groupement et la position qui donnent la somme la plus élevée (ex. le groupement qui correspond à la forme masculine avec consonne de liaison, et cela pour la position 2). A la fin de la comparaison, les informations concernant le groupement et la position sont stockées avec le mot. Le processus recommence à chaque nouveau cycle et peut donc produire un nouveau résultat, le processeur de position ne retenant qu'un seul groupement et qu'une seule position pour chaque mot. Notons que des informations numériques peuvent être attachées aux groupements de connexions afin de rendre compte de la préférence, ou de la fréquence, d'un groupement donné chez l'être humain (voir les mots avec schwa facultatif, par exemple, où il y a souvent une préférence pour l'une ou l'autre version).

3) *Point d'isolation*

Dans la tâche expérimentale du 'gating' (Grosjean, 1980), les sujets entendent un mot, présenté en segments de plus en plus longs, du début à la fin. Lors de la première présentation, ils n'entendent que le segment initial du mot (ex. les premières 30 ms); lors de la deuxième présentation, ils reçoivent un segment qui est le double en longueur (ex. les premières 60 ms), et ainsi de suite, jusqu'à ce qu'ils aient entendu, lors de la dernière présentation, le mot en entier. Après chaque présentation, les sujets proposent le mot en question en l'énonçant ou en l'écrivant sur papier. En examinant les réponses, le chercheur peut établir le point d'isolation du mot pour chaque participant, à savoir le point qui correspond à la durée du segment, dans la séquence de présentations, pour laquelle il a indiqué le mot en question et n'a plus modifié sa réponse par la suite.

Nous avons adapté cette mesure pour l'utiliser dans le modèle et ce de la façon suivante. A chaque cycle de la simulation, le modèle établit les niveaux d'activation de tous les mots, en se basant à la fois sur les signaux ascendants et latéraux vers les mots, et sur leur fréquence d'occurrence. Les mots qui ressemblent à l'entrée accumulent de l'activation, d'un cycle à l'autre. Quant au mot cible (le mot à reconnaître), son niveau d'activation augmente encore plus rapidement (dans le cas d'un déroulement normal de la simulation) car il correspond parfaitement à la chaîne de phonèmes en entrée. Le point d'isolation (PI) défini par le modèle est le moment, exprimé en nombre de cycles, où le mot cible a dépassé le niveau d'activation de tous les autres candidats possibles; de plus, ce premier rang doit être maintenu jusqu'à la fin de la simulation. Cette mesure est proposée en nombre de cycles ainsi qu'en pourcentage de la longueur du mot (neutralisant ainsi la longueur). Même s'il reflète une réalité légèrement différente de la variante expérimentale, le PI du modèle est une mesure fort utile dans la simulation de la reconnaissance de mots parlés.

2 EXEMPLES DE SIMULATION

Dans ce qui suit, nous présentons des exemples de simulation de la reconnaissance de différents types de mot, ou suite de deux mots. Pour ce faire, nous avons utilisé

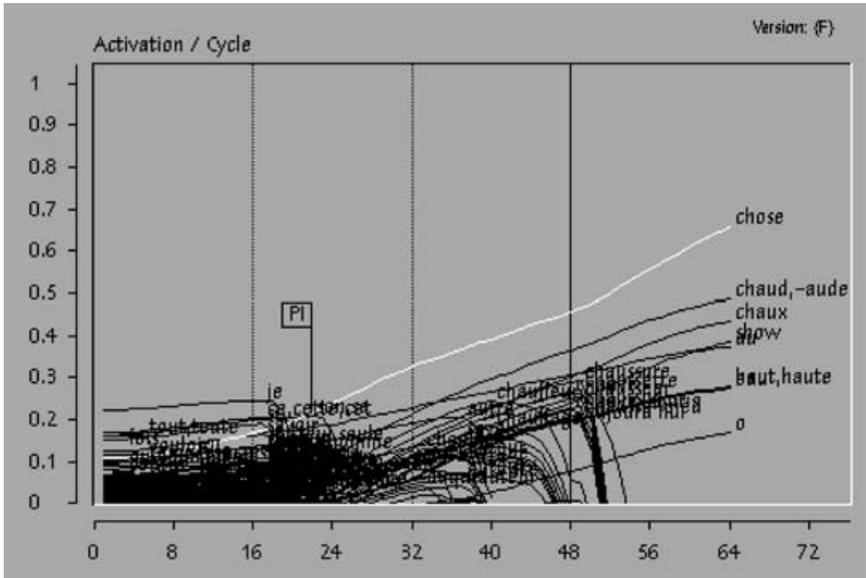


Figure 5. Simulation de la reconnaissance de ‘chose’ (/ʒoz/), mot monosyllabique de haute fréquence.

un lexique complet, dans sa version française, sauf pour un exemple où un lexique Suisse romand a été utilisé. Pour chaque exemple, nous avons entré la chaîne phonétique entière, phonème par phonème, sans arrêter la simulation après chaque phonème (bouton ‘Entrer tout’), et nous montrons l’état d’activation des mots à la fin du processus. Nous ne verrons donc pas les étapes intermédiaires mais nous y ferons référence lorsque cela s’avérera utile.

Mot monosyllabique de haute fréquence

La Figure 5 montre la simulation de la reconnaissance du mot ‘chose’, mot de trois phonèmes et d’une fréquence d’occurrence élevée (0.712; valeur normalisée). Rappelons que dans la littérature psycholinguistique, l’effet fréquence est l’un des effets les plus solides qui soit. Lors des 16 premiers cycles au cours desquels le /ʒ/ initial de ‘chose’ est présenté au modèle⁴, plusieurs mots potentiels sont temporairement activés. Parmi ceux-ci se trouvent des candidats qui commencent avec le phonème /ʒ/ (ex. ‘chose’; ‘chaque’; ‘chambre’, etc.) ou avec un phonème assez proche, comme /s/ ou /ʃ/ (ex. ‘je’; ‘si’, etc.). De plus, nous trouvons des candidats qui sont très fréquents (ex. ‘tout, toute’). Les mots les plus actifs sont

⁴ Le nombre de cycles par phonème correspond au nombre de fois que la simulation exécute, pour le phonème en question, l’ensemble des processus internes du modèle. Plus ce nombre est élevé, plus la simulation est fine, et plus elle se rapproche de l’idéal théorique d’une simulation sans étapes discrètes; mais le désavantage est alors que la simulation est plus coûteuse en temps. La valeur de 16 cycles s’avère être un bon compromis.

Homophones en français standard mais pas en français de Suisse romande

Certaines paires de mots sont homophones en français standard mais pas en français de Suisse romande comme, par exemple, ‘ami’ et ‘amie’ qui se prononcent tous deux /ami/ en France, mais /ami/ (/i/ court) et /ami:/ (/i:/ long) en Romandie. Quel est le comportement du modèle lorsque l’on entre ‘ami’ en se servant du lexique français (qui ne contient qu’un seul son /i/) ou du lexique Suisse romand (qui contient deux phonèmes)? Comme nous l’observons dans la Figure 8 (haut), il y a un cas d’homophonie typique dans la simulation en français standard: ‘ami’ et ‘amie’ sont tous deux actifs et, puisqu’ils ont la même fréquence dans le modèle, ils partagent une seule et même courbe d’activation.⁵ Par contre, dans la simulation en français de Suisse romand (bas de la figure), nous voyons que les deux mots se séparent légèrement pendant le /i/ de ‘ami’ (cycles 33–48). Ensuite, pendant le silence de fin de mot (cycles 49–64), ‘ami’ continue à monter alors que ‘amie’ chute assez rapidement. Notons une autre différence entre les deux simulations. Les concurrents au deuxième rang (qui jouent un rôle capital pour le PI du mot cible) sont dissemblables dans les deux cas: ‘âme’ en français standard (haut de la figure) et ‘amie’ en français de Suisse romande (bas de la figure), les PI du mot cible étant à 42 et 37 cycles respectivement. Cela vient du fait que ‘âme’ français (/am/) fait partie de ‘ami’ (/ami/), tandis que ‘âme’ Suisse romand commence avec une autre voyelle (/am/) et n’est donc pas un candidat sérieux. Bien entendu, et comme il se doit, les autres mots activés sont très similaires dans les deux cas.⁶

Mot avec schwa

Pour illustrer les lexèmes qui possèdent plusieurs formes phonétiques, nous avons choisi les mots qui contiennent des schwas facultatifs et qui peuvent donc être prononcés avec ou sans l’effacement de cette voyelle.⁷ Dans la Figure 9, nous montrons le parcours de reconnaissance de ‘pelouse’ sans effacement du schwa (/pøluz/) (haut de la figure), et avec effacement (/pluz/) (bas de la figure). En comparant les deux simulations, nous remarquons que les premiers candidats, activés pendant l’entrée du /p/ (cycles 1–16), sont identiques, comme il se doit (ex. ‘mais’;

⁵ Bien entendu, le modèle ne contient aucune information sur l’accord grammatical, le contexte, ou le sens des mots, et il n’est donc pas capable de préférer l’un ou l’autre mot pour ces raisons. Ceci dit, il n’a apparemment pas de difficulté à activer des homophones bien que ceux-ci s’inhibent mutuellement.

⁶ Notons la manière dont le modèle se comporte pour les homophones qui n’ont pas la même fréquence, cas que l’on trouve avec ‘chat’, ‘chas’ et ‘schah’, par exemple. Bien que ces trois mots soient activés lorsque l’on présente la chaîne /ʃa/, le mot ‘chat’ dépasse immédiatement les deux autres car il est le plus fréquent. Ces derniers restent actifs jusqu’à la fin de la simulation, mais toujours à un niveau inférieur à ‘chat’.

⁷ Rappelons que le modèle simule également la reconnaissance d’adjectifs qui possèdent plusieurs formes phonétiques (ex. ‘grand(e)’ avec ces trois formes: masculin sans liaison, masculin avec liaison, et féminin). Le mécanisme sous-jacent, qui se sert des groupements de connexions, est identique.

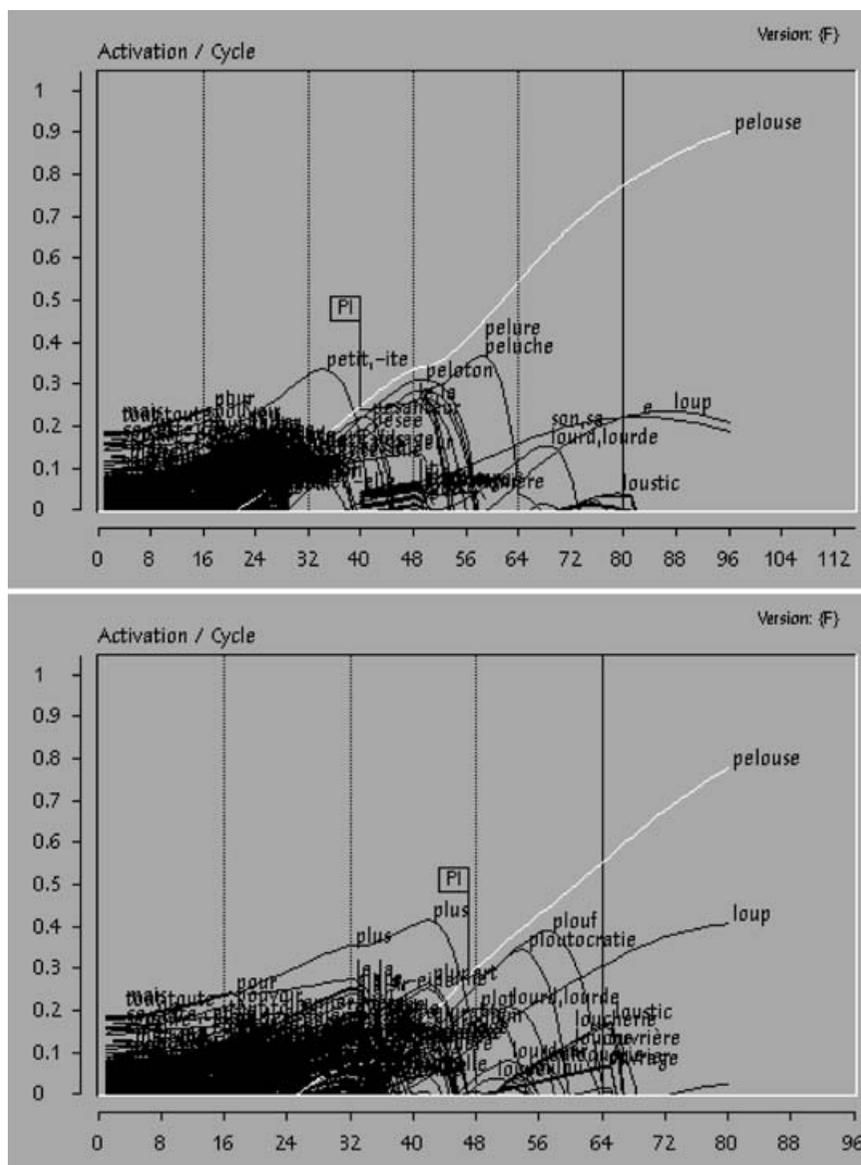


Figure 9. Simulation de la reconnaissance de ‘pelouse’, sans effacement du schwa (/pəluz/) (haut de la figure), et du même mot mais présenté avec effacement du schwa (/pluz/) (bas de la figure).

figure) active surtout des mots qui débutent avec /pl/ (ex. ‘plus’; ‘place’; ‘plaisir’; ‘plupart’, etc.). Les concurrents sont également moins nombreux et subsistent moins longtemps dans le haut de la figure ce qui est dû à un voisinage lexical moins dense

pour /pølu:/ que pour /pluz/. Dans la dernière étape des deux simulations (à partir du cycle 64), les candidats sont à nouveau presque identiques (ex. 'loup'; 'lourd, lourde'; 'loustic', etc.), le mot 'pelouse' dépassant ceux-ci, sans difficulté, et ce dans les deux cas. Il est intéressant de noter que le PI se situe au cycle 40 pour /pølu:/ et au cycle 47 pour /pluz/ ce qui simule correctement la reconnaissance plus précoce de la version avec schwa chez l'auditeur humain (Racine et Grosjean, 2000). Notons enfin que le modèle est capable de simuler une reconnaissance différente des mots avec effacement obligatoire du schwa (ex. 'médecin'), effacement facultatif (ex. 'pelouse'), et effacement interdit (ex. 'vendredi').

Deux mots, séparés ou non par une pause

Le modèle permet de simuler la reconnaissance de mots isolés mais également de suites de deux mots, ce qui est encore rare dans les modèles actuels. Prenons tout d'abord le cas de deux mots séparés ou non par une pause. Dans la Figure 10 (haut), nous présentons le résultat final de la reconnaissance simulée de 'sa', suivi d'une pause, et ensuite de 'voix' (/sa-vwa/), et dans la même figure (bas), nous proposons ces deux mots sans pause intermédiaire (/savwa/). Dans le cas de la séparation avec pause (celle-ci correspond à la troisième colonne à partir de la gauche dans le haut de la figure), on trouve deux nuages d'activation distincts (l'un pour 'sa' et l'autre pour 'voix'), deux courbes différentes pour les mots cibles (à noter que 'son, sa' est un déterminant avec plusieurs formes et que 'voix' est un homophone) ainsi que deux valeurs de PI bien séparées l'une de l'autre ('sa' est isolé au cycle 23 et 'voix' au cycle 70). On constate aussi que 'son, sa' continue à monter pendant l'entrée du mot 'voix', et que les deux mots sont actifs, en occupant respectivement le premier et le deuxième rang, à la fin de la simulation. Le modèle traite donc correctement la reconnaissance de ces deux mots.

Ce qui se passe lorsque ces deux mêmes mots ne sont pas séparés par une pause est assez différent (voir le bas de la figure) car le modèle ne reçoit aucune information concernant la fin de 'sa' et le début de 'voix'. On remarque d'ailleurs qu'il n'y a qu'un seul nuage d'activation de mots. Pendant l'entrée du /s/ et du /a/ (cycles 1-32), les candidats sont les mêmes que ceux du haut de la figure, mais pendant l'entrée du /v/, du /w/ et du /a/ de 'voix' (cycles 33-80), au lieu que ce soit 'voix' qui s'active, on observe que ce mot est inhibé par des mots qui débudent avec /sav/ et qui sont eux-mêmes des candidats (ex. 'savoir'; 'savant, -ante'; 'saveur', etc.). Cela aboutit à ce que 'savoir' dépasse même 'son, sa' au cycle 58 et devienne le candidat le plus actif alors que 'savoir-faire' et 'savoir-vivre' sont d'excellents compétiteurs proches. Nous avons clairement ici un cas d'impasse. Ce n'est que pendant le silence qui marque la fin de 'voix' (cycles 81-96) que les trois compétiteurs s'avèrent inadéquats et chutent nettement. 'Son, sa' reprend le premier rang et 'voix' passe au deuxième rang. Les deux sont isolés à peu près au même moment (aux cycles 84 et 86 respectivement) mais beaucoup plus tardivement que ces deux mêmes mots dans le haut de la Figure 10. Il est clair qu'une marque de frontière de mot (ex. une pause) facilite grandement la reconnaissance lexicale d'une suite de deux mots.

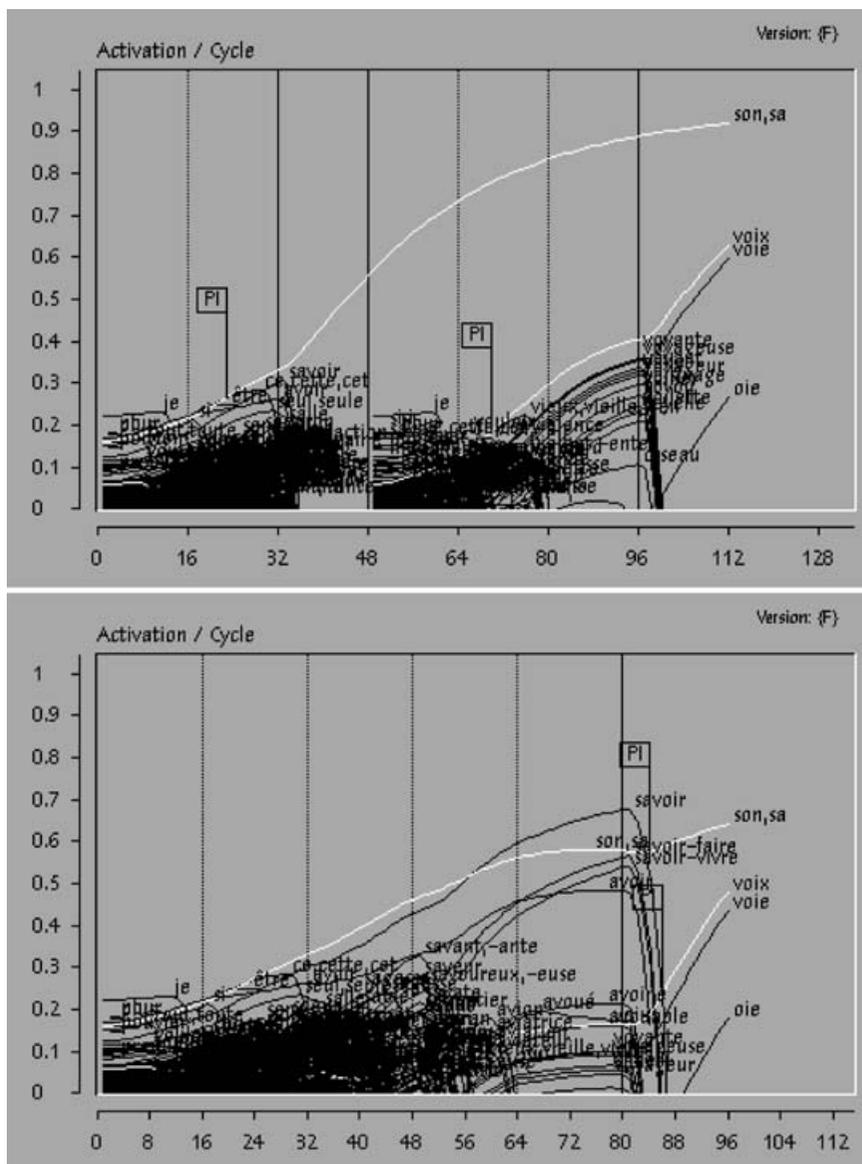


Figure 10. Simulation de la reconnaissance de 'sa voix' où les mots sont séparés par une pause (/sa-vwa/) (haut de la figure) et où ils ne sont pas séparés par une pause (/savwa/) (bas de la figure).

La liaison

Enfin, le modèle peut traiter les mots enchaînés avec liaison. La Figure 11 reprend l'exemple de 'petit ami', /pitami/, transcrit avec effacement du schwa dans 'petit'

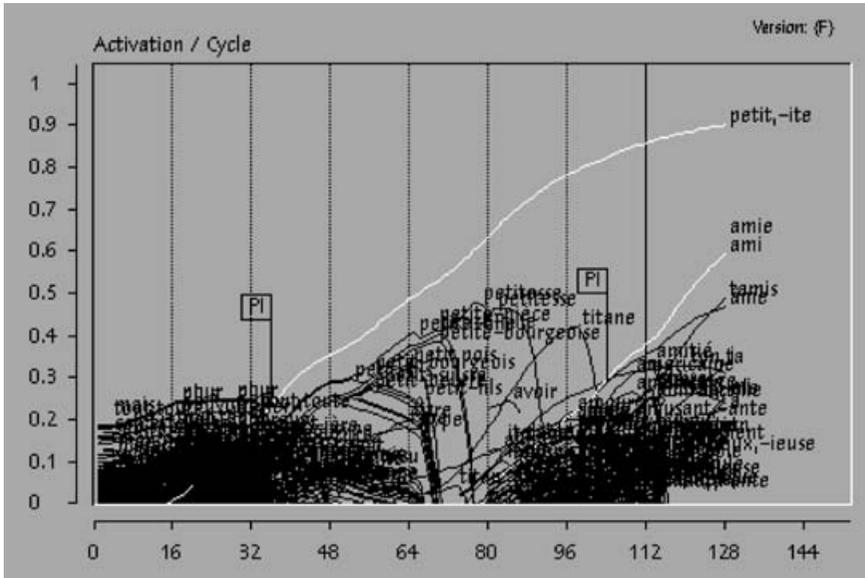


Figure 11. Simulation de la reconnaissance complète de ‘petit ami’ (/ptitami/).

et avec une consonne de liaison /t/ avant ‘ami’. Contrairement à la Figure 2, nous montrons ici le résultat complet de la simulation (voir la discussion plus haut pour la simulation de la reconnaissance de ‘petit, -ite’ dont le PI est au cycle 36). Pendant l’entrée du /t/ de liaison (cycles 49–64), ‘petit, -ite’ continue à monter tandis que des mots non compatibles (ex. les mots composés du type ‘petit-fils’) retombent. Quand le /a/ de ‘ami’ arrive (cycles 65–80), quelques nouveaux candidats entrent en jeu (ex. ‘titane’; ‘italien’; ‘italienne’, etc.). Pendant l’entrée du /m/ (cycles 81–96), les mots en /am/ (ex. ‘ami’; ‘amour’; ‘amiral’, etc.) se renforcent, et c’est pendant la dernière voyelle et la marque de fin de suite (cycles 97–128) que ‘ami’, couplé à l’homophone ‘amie’, est isolé (PI au cycle 104). Notons que la chaîne /ptitami/ aurait aussi pu être reconnue par le modèle comme étant ‘petit tamis’, et donc sans liaison. En effet, ‘tamis’ est assez actif à la fin de la simulation, mais comme ce mot est moins fréquent que ‘ami’, il se trouve à un niveau d’activation inférieur.

3 AUTRES ASPECTS

Dans cette troisième partie, nous donnons des détails sur quelques autres aspects du modèle.

Modélisation du traitement simultané

Bien que le signal physique de la parole soit séquentiel, les chercheurs en psycholinguistique ne sont pas tous d’accord pour affirmer que la reconnaissance

de mots parlés s'effectue d'une manière purement séquentielle, elle aussi. En effet, à côté de l'approche séquentielle 'standard' (ex. Marslen-Wilson et Welsh, 1978; McClelland et Elman, 1986), plusieurs chercheurs ont proposé, dans des modèles verbaux, une approche plus ou moins simultanée (ex. Grosjean et Gee, 1987; Christophe et Dupoux, 1996; Wauquier-Gravelines, 1996, etc.). Un modèle informatique qui incorpore les deux approches (séquentielle et simultanée) est donc souhaitable afin d'étudier les avantages et les inconvénients de l'une et de l'autre. C'est en faisant varier la valeur du délai d'entrée des phonèmes que nous permettons à l'utilisateur de naviguer tout au long du continuum séquentiel – simultané. Le mode standard reste le traitement entièrement séquentiel, mode dans lequel les phonèmes arrivent les uns après les autres, du début à la fin du mot. Il est obtenu en utilisant un délai d'entrée de 16 cycles (rappelons que la durée d'entrée des phonèmes est de 16 cycles, elle aussi). Nous pouvons également choisir un mode de traitement à mi-chemin entre le séquentiel et le simultané (en utilisant un délai d'entrée de 8 cycles), ou encore avoir un mode de traitement entièrement simultané (délai d'entrée de 0 cycle). Ce mode n'est qu'hypothétique pour les mots longs mais il est plausible pour les mots courts.

La question de l'inhibition ascendante des phonèmes vers les mots

L'inhibition ascendante est présente dans la plupart des modèles de reconnaissance de mots parlés et elle peut être soit catégorique, lorsqu'il y a élimination immédiate d'un candidat activé dès que la dissemblance est détectée (ex. Marslen-Wilson et Welsh, 1978), soit graduelle, lorsque le niveau d'activation du mot en question est simplement réduit (ex. Marslen-Wilson, 1987; Norris, 1994). Le modèle TRACE est l'un des seuls où il n'y a pas d'inhibition ascendante car ses auteurs (McClelland et Elman, 1986) postulent qu'elle est superflue, le même effet pouvant être obtenu à l'aide de l'inhibition latérale entre mots. Bien qu'il soit difficile de tester cette proposition expérimentalement, Frauenfelder, Scholten et Content (2001) ont avancé dans cette direction, et ont confirmé l'importance de l'inhibition ascendante dans un modèle humain. Nous l'avons donc implémentée dans le nôtre.

La question de la rétroaction descendante des mots vers les phonèmes

Le débat dans la littérature reste vif sur la question de la rétroaction des mots vers les phonèmes. Il y a ceux qui la défendent (ex. Elman et McClelland, 1988; Magnuson, McMurray, Tanenhaus et Aslin, 2003; Samuel et Pitt, 2003) et d'autres qui s'y opposent (ex. Massaro, 1989; Norris, McQueen et Cutler, 2000). C'est précisément parce qu'il existe cette incertitude dans la littérature que les liens descendants mots-phonèmes sont optionnels dans notre modèle (voir la Figure 4, plus haut). Normalement, les deux mécanismes concernés, l'activation et l'inhibition descendantes, sont mis à zéro, mais nous gardons la possibilité de les enclencher lorsque cela est souhaité. De cette façon, le modèle permet à son utilisateur d'étudier une question complexe, et encore sujette à bien des débats.

ÉTAPES SUIVANTES

Nous souhaitons étendre le comportement du modèle dans deux directions principales: la coupure lexicale entre mots, et le traitement des élisions. En ce qui concerne le premier point, le modèle simule actuellement la reconnaissance de suites de deux mots présentées sans pause entre les deux mots, ou avec pause. Celle-ci couvre à la fois la pause silencieuse et la pause remplie, et sert de coupure lexicale nette. Or, il existe d'autres indices de coupure lexicale dans la parole, notamment l'allongement de la fin du mot qui précède celle-ci (Monnin et Grosjean, 1993; Quené, 1992). Notre intention est de permettre au modèle de repérer l'allongement de la consonne finale du premier mot dans les suites de deux mots. Il devrait alors montrer plus de facilité à reconnaître 'chaque ours' lorsqu'il est entré avec une consonne allongée (/ʃak:uʁs/) au lieu d'hésiter entre 'chaque ours' et 'chaque course' comme cela se passe actuellement. En ce qui concerne le second point, les élisions, le modèle actuel ne traite que l'effacement du schwa (ex. 'p'louse') et nous souhaiterions étendre son comportement à d'autres types d'élisions (ex. 'l'arbre', 'c'matin', 'c't après-midi', 'vot' tour', etc.). Certes, certaines d'entre elles ne se font pas au niveau orthographique, mais elles sont fréquentes au niveau oral, la modalité sur laquelle nous travaillons.

Outre la simulation d'effets bien connus en reconnaissance de mots parlés, et l'évaluation systématique du modèle à l'aide d'études paramétriques, nous utiliserons celui-ci pour comprendre certains résultats expérimentaux difficiles à expliquer. Par exemple, Wauquier-Gravelines (1996) a testé 'un petit éléphant' (liaison) par rapport à 'un joli éléphant' (pas de liaison) avec une tâche de détection de mot, les sujets devant détecter le mot 'éléphant', et elle n'a trouvé aucune différence significative. Or, on aurait pu s'attendre, dans le cas de la liaison, à une compétition entre les candidats commençant par la consonne de liaison et ceux commençant par une voyelle, ce qui aurait eu pour conséquence de retarder la reconnaissance du substantif. Une explication possible est qu'il s'agit d'un effet très éphémère qui disparaît avant la fin du mot et qui ne peut être montré que difficilement dans une étude expérimentale. A l'aide du modèle, nous pourrions observer s'il y a bien une ambiguïté momentanée, et donc une compétition entre les candidats commençant par une consonne et ceux commençant par une voyelle. Si cela devait être le cas, cela encouragerait l'expérimentateur à trouver une tâche qui soit sensible à cette ambiguïté très momentanée. De cette manière, le modèle contribuera au va-et-vient entre recherche expérimentale et théorisation.

Adresse pour la correspondance:

Nicolas Léwy

Laboratoire de traitement du langage et de la parole

Université de Neuchâtel

Avenue du Premier-Mars 26

CH-2000 Neuchâtel

Suisse

e-mail: nicolas.lewy@unine.ch

REMERCIEMENTS

Ce projet a été rendu possible grâce à un subside du Fonds national suisse de la recherche scientifique (FN 1214-058848.99). Le modèle est une extension de la partie française de BIMOLA, modèle implémenté par le premier auteur. Nous tenons à exprimer notre reconnaissance à Jeff Elman (Center for Research in Language, UC San Diego) pour ses conseils très précieux, et à Rachel Heim pour son aide lors de l'entrée des données.

RÉFÉRENCES

- Ans, B., Carbonnel, S. et Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, 105(4): 678–723.
- Bagley, W. (1900). The apperception of the spoken sentence: a study in the psychology of language. *American Journal of Psychology*, 12(1): 80–130.
- Christiansen, M., Allen, J. et Seidenberg, M. (1998). Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes*, 13(2/3): 221–268.
- Christophe, A. et Dupoux, E. (1996). Bootstrapping lexical acquisition: the role of prosodic structure. *Linguistic Review*, 13(3/4): 383–412.
- Content, A., Mousty, P. et Radeau, M. (1990). Brulex: une base de données lexicales informatisée pour le français écrit et parlé. *L'Année psychologique*, 90(4): 551–566.
- Crocker, M. (1996). *Computational psycholinguistics: an interdisciplinary approach to the study of language*. Dordrecht, Pays-Bas: Kluwer Academic.
- Dijkstra, T. et De Smedt, K. (eds). (1996). *Computational psycholinguistics: AI and connectionist models of human language processing*. Londres: Taylor & Francis.
- Elman, J. et McClelland, J. (1988). Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27(2): 143–165.
- Frauenfelder, U. et Peeters, G. (1998). Simulating the time course of word recognition: an analysis of lexical competition in TRACE. Dans: J. Grainger et A. Jacobs (eds), *Localist connectionist approaches to human cognition*. Mahwah, NJ: Erlbaum, pp. 101–146.
- Frauenfelder, U., Scholten, M. et Content, A. (2001). Bottom-up inhibition in lexical selection: phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, 16(5/6): 583–607.
- Gaskell, G., Hare, M. et Marslen-Wilson, W. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, 19(4): 407–439.
- Grainger, J. et Jacobs, A. (eds). (1998). *Localist connectionist approaches to human cognition*. Mahwah, NJ: Erlbaum.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28(4): 267–283.
- Grosjean, F. et Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25(1/2): 135–155.
- Luce, P., Goldinger, S., Auer, E. et Vitevitch, M. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, 62(3): 615–625.
- Magnuson, J., McMurray, B., Tanenhaus, M. et Aslin, R. (2003). Lexical effects on compensation for coarticulation: the ghost of Christmas past. *Cognitive Science*, 27(2): 285–298.

- Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1/2): 71–102.
- Marslen-Wilson, W. et Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1): 29–63.
- Massaro, D. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, 21(3): 398–421.
- McClelland, J. et Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1): 1–86.
- McClelland, J. et Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: part 1. An account of basic findings. *Psychological Review*, 88(5): 375–407.
- Métral, J.-P. (1977). Le vocalisme du français en Suisse romande: considérations phonologiques. *Cahiers Ferdinand de Saussure*, 31: 145–176.
- Monnin, P. et Grosjean, F. (1993). Les structures de performance en français: caractérisation et prédiction. *L'Année psychologique*, 93(1): 9–30.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76(2): 165–178.
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52(3): 189–234.
- Norris, D., McQueen, J. et Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, 23(3): 299–325.
- Page, M. (2000). Connectionist modelling in psychology: a localist manifesto. *Behavioral and Brain Sciences*, 23(4): 443–467.
- Plunkett, K. et Elman, J. (1997). *Exercises in rethinking innateness: a handbook for connectionist simulations*. Cambridge, MA: MIT Press.
- Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20(3): 331–350.
- Racine, I. et Grosjean, F. (2000). Influence de l'effacement du schwa sur la reconnaissance des mots en parole continue. *L'Année psychologique*, 100(3): 393–417.
- Samuel, A. et Pitt, M. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48(2): 416–434.
- Wauquier-Gravelines, S. (1996). *Organisation phonologique et traitement de la parole continue, contraintes prosodiques et phonologiques de l'accès lexical*. Thèse de doctorat de linguistique théorique et formelle, Université Paris VII.