

PREMIÈRE PARTIE

Modèles et composantes
de la production verbale

Chapitre 1

Les modèles de la production de la parole

1.1. Introduction

« Parler est une de nos préoccupations préférées. Nous passons plusieurs heures par jour à discuter, raconter des histoires, débattre, argumenter... et, bien sûr, à nous parler à nous-mêmes » (Levelt, 1989). Au cours d'une conversation normale, nous produisons 2 à 3 mots par seconde, ce qui correspond à environ 4 syllabes et 10-12 phonèmes par seconde. Ces mots sont sélectionnés extrêmement rapidement (en moins d'une seconde) dans le lexique mental qui comprend environ 60 000 mots chez un adulte lettré. Nous faisons très peu d'erreurs de production, moins d'une erreur sur 1 000 mots produits (Rossi et Peter-Defare, 1998). Nous sommes nés pour parler, et parvenus à l'âge adulte, nous avons produit environ 50 millions de mots. C'est sans doute le talent cognitif et moteur que nous pratiquons le plus ; c'est celui qui nous paraît le plus simple et le plus immédiat, mais c'est aussi le plus complexe.

1.2. La production de la parole : trois étapes principales

La plupart des théories psycholinguistiques (Dell, 1986; Levelt, 1989) suggèrent que la production de la parole passe par trois étapes principales de traitement (voir figure 1.1) : (1) la première étape est celle de la *préparation conceptuelle* du message préverbal (le message à exprimer correspond aux intentions, aux idées que le locuteur veut exprimer ; ce message est non linguistique ou préverbal). Ce message contient un ou plusieurs concepts pour lesquels nous avons des mots dans notre lexique mental et ces mots doivent être récupérés ; (2) la seconde étape est celle de la *formulation* ou *lexicalisation* du message préverbal : cette étape d'accès

au lexique est elle-même divisée en deux sous-étapes; il s'agit tout d'abord de récupérer les informations sémantiques et syntaxiques relatives aux mots concernés (*étape de sélection lexicale*), il faut ensuite récupérer les informations phonologiques et morphologiques correspondantes (*étape d'encodage phonologique*) ; (3) la troisième étape est celle de l'*articulation*.

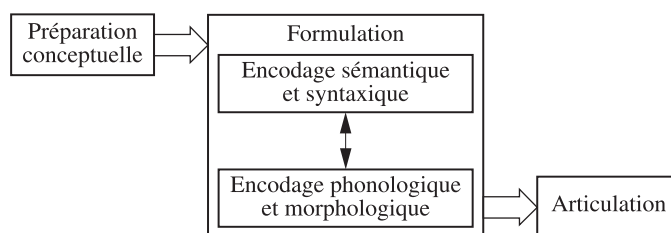


Figure 1.1. Les trois étapes principales impliquées dans la production de la parole : (1) conceptualisation ; (2) formulation (encodage sémantique/syntaxique et phonologique/morphologique) ; (3) articulation

Dans ce chapitre, nous traiterons essentiellement de la seconde étape, celle de la *formulation* ou de la *lexicalisation* dans la mesure où elle concerne l'accès au lexique au cours de la production. De nombreux travaux suggèrent l'existence de deux niveaux de traitement indépendants mais interconnectés mis en œuvre dans l'accès au lexique au cours de la production de parole (Ferrand, 1994, 2001a, 2001b ; Segui et Ferrand, 2000). Les données issues des erreurs de production (Rossi et Peter-Defare, 1998), du phénomène du mot sur le bout de la langue (Vigliocco, Antonini et Garrett, 1997), de la chronométrie mentale (Schriefers, Meyer et Levelt, 1990; Levelt, Schriefers, Vorberg, Meyer et Pechmann, 1991), de la neuropsychologie cognitive (Caramazza, 1997) et de l'imagerie cérébrale conduisent toutes à ce découpage en deux niveaux de traitements : (1) le niveau sémantique/syntaxique (sélection lexicale), (2) le niveau d'encodage phonologique, correspondant respectivement à la récupération des lemmas (informations sémantiques et syntaxiques) et des lexèmes (informations phonologiques et morphologiques).

Toutefois, les modèles diffèrent sur tous les autres points (Ferrand, 2001a, 2001b ; Segui et Ferrand, 2000). Les divergences portent en particulier sur la nature des représentations des différents niveaux concernés :

- les représentations sont-elles locales (Dell, 1986 ; Dell, Schwartz, Martin et Gagnon, 1997) ou distribuées (Dell, Juliano et Govindjee, 1993) ?
- l'information conceptuelle est-elle représentée sous la forme de traits sémantiques décomposables (Caramazza, 1997 ; Dell, 1986) ou bien est-elle globale et indivisible (Roelofs, 1997a) ?

– les représentations morphologiques sont-elles décomposables (Roelofs, 1996) ou non ?

– la récupération des informations syntaxiques est-elle obligatoire (Levelt, Roelofs et Meyer, 1999) ou optionnelle (Caramazza, 1997 ; Caramazza et Miozzo, 1997) ?

– existe-t-il ou non des connexions inhibitrices à l'intérieur de chaque niveau et entre les différents niveaux (comme cela est suggéré par Berg et Schade, 1992 ; et contrairement à ce que proposent Levelt *et al.*, 1999) ?

Les divergences entre les différents modèles portent également sur le déroulement temporel de ces deux étapes d'encodage sémantique/syntaxique et d'encodage phonologique/morphologique : ces deux étapes de lexicalisation sont-elles strictement sérielles (Levelt *et al.*, 1991, 1999), en cascade (Peterson et Savoy, 1998), ou interactives (Dell, 1986 ; Dell *et al.*, 1997 ; Cutting et Ferreira, 1999 ; Rapp et Goldrick, 1999) ?

1.3. Les principales techniques utilisées pour étudier la production de la parole

Classiquement, l'étude des mécanismes impliqués dans la production de la parole repose sur deux approches différentes : la première, la plus ancienne, consiste à recueillir et à analyser les erreurs spontanées de production ou à les induire en laboratoire ; la seconde approche est celle de la chronométrie mentale, qui consiste à enregistrer et à analyser les latences au cours de la production de mots ou de la dénomination de dessins d'objets. Nous verrons dans les paragraphes suivants que ces deux approches, bien que n'étant pas incompatibles, ont conduit au développement de deux modèles opposés de la production de la parole : le modèle interactif de Dell (1986) et le modèle sériel de Levelt *et al.* (1991).

Jusqu'à très récemment, presque toutes les recherches sur la production de la parole reposaient sur l'analyse des erreurs produites lors de conversations de la vie de tous les jours. Ces recherches ont permis d'identifier les principaux types d'erreurs à partir desquelles ont été élaborées des hypothèses largement partagées. En particulier, ces hypothèses ont été formulées à partir de la prise en considération de la nature régulière des erreurs, c'est-à-dire du fait que celles-ci semblent obéir à des règles. Les erreurs de production ont été depuis longtemps considérées comme fournissant des informations précieuses quant aux mécanismes cognitifs mis en œuvre au cours de la production de la parole. En particulier, la nature des erreurs peut indiquer à quel niveau le système est pris en défaut. Ceci permet d'associer chaque type d'erreurs à une étape particulière de traitement. Leur analyse peut donc permettre de faire des inférences sur l'architecture fonctionnelle du système de production, c'est-à-dire sur son organisation interne. Il existe d'importants recueils d'erreurs de production en anglais (Fromkin, 1973, par exemple) et, plus récemment, en français (Rossi et Peter-

Defare, 1998). Les erreurs de production prennent des formes multiples. Toutefois, dans cette diversité, une taxinomie précise a été élaborée (Dell, 1986). L'observation des erreurs de production présente un intérêt considérable mais aussi de nombreux inconvénients, notamment la rareté des erreurs, l'ambiguïté de leur nature, ainsi que les biais dans leur recueil. Les erreurs de syntaxe surviennent moins de 5 fois sur 1 000 phrases produites, les erreurs de sélection lexicale surviennent moins d'une fois sur 1 000 mots produits, et les erreurs d'encodage phonologique moins de 4 fois sur 10 000 mots produits. Néanmoins, malgré tous ces problèmes potentiels, il existe une bonne convergence entre les observations naturelles et celles induites en laboratoire.

Une autre méthode très fréquemment utilisée pour étudier la production de la parole est celle de la dénomination d'objets (Ferrand, 1997). Cette technique permet d'étudier le décours temporel des processus impliqués dans la production de la parole. L'analyse des temps de réaction constitue en psychologie expérimentale l'un des instruments privilégiés pour inférer la complexité et la nature des processus étudiés. Que se passe-t-il quand nous devons dénommer un dessin d'objet ? Généralement, cela nous prend entre 600 et 1 200 millisecondes en moyenne entre la présentation du dessin et l'initialisation de la réponse verbale, et nous commettons très peu d'erreurs. Dénommer un objet demande plusieurs étapes de traitement, qui peuvent ou non se recouvrir temporellement, comme nous le verrons par la suite. Premièrement, le locuteur doit traiter le dessin visuellement. Cela implique l'extraction de traits visuels : bords, coins, ombres, etc. Bien sûr, les étapes initiales de la dénomination, celles qui sont impliquées dans l'analyse et la reconnaissance visuelle de l'objet, ne font pas partie des processus de production de la parole proprement dit.

Deuxièmement, le locuteur doit catégoriser le dessin comme étant une maison, un avion, une montre ou autre. Cette étape lui permet de recruter le mot sémantiquement approprié dans le lexique mental. Cela lui fournit de l'information sur la façon dont le mot représenté est utilisé habituellement : s'agit-il d'un verbe, d'un nom ou autre ? Est-il au masculin ou féminin, au singulier ou pluriel, etc. ? Localiser une entrée lexicale exige qu'un mot adéquat existe dans le lexique mental du locuteur pour véhiculer le message correspondant. La facilité de sélection d'un mot pour la production dépend de nombreux facteurs comme, par exemple, la fréquence du nom de l'objet (plus un mot est fréquent ou familier, et plus vite a lieu sa récupération), l'âge d'acquisition du nom de l'objet (plus un mot est acquis tôt, et plus vite a lieu sa récupération), etc. Une fois l'entrée lexicale sélectionnée, la prononciation du mot doit être récupérée et produite.

1.4. Les principaux modèles de la production de la parole

Nous présentons trois principaux modèles de la production de la parole (voir Ferrand, 2001b pour plus de détails) : le modèle en cascade avec interactivité de

Dell (Dell, 1986 ; Dell *et al.*, 1997), le modèle strictement sériel de Levelt (Levelt *et al.*, 1999) et le modèle à réseau indépendant de Caramazza (Caramazza, 1997 ; Caramazza et Miozzo, 1997). Ces trois modèles sont « localistes » avec des représentations mentales symboliques discrètes comme des traits sémantiques, des phonèmes, des morphèmes, etc. (le seul modèle distribué étant celui de Dell *et al.*, 1993).

1.4.1. *Le modèle en cascade interactif de Dell*

Gary Dell (Dell, 1986 ; Dell *et al.*, 1997) a proposé un modèle connexionniste en cascade et interactif à deux étapes comprenant trois niveaux de représentations (voir figure 1.2) : il s'agit du niveau sémantique (correspondant à des traits sémantiques), du niveau lexical (correspondant aux mots) et du niveau phonologique (correspondant aux phonèmes). Ces trois niveaux sont reliés par des connexions excitatrices bidirectionnelles (il n'existe pas de connexions inhibitrices dans ce modèle). De plus, chaque niveau interagit avec les niveaux adjacents. Les connexions descendantes vont du niveau sémantique vers le niveau lexical, puis du niveau lexical vers le niveau phonologique. Les connexions ascendantes vont dans l'autre sens, permettant l'interactivité du modèle grâce à cette rétroaction d'activation du niveau phonologique vers le niveau lexical et sémantique. De ce point de vue, ce modèle est similaire aux autres modèles interactifs de la production de la parole (Stemberger, 1985) et diffère des modèles strictement sériels (Levelt *et al.*, 1991, 1999). Les représentations de chaque niveau sont codées par des valeurs d'activation. Ces valeurs sont actualisées au cours du traitement en fonction des niveaux d'activation des représentations voisines, le déclin de l'activation de base, et l'influence du bruit aléatoire. Les mots sont récupérés grâce à la propagation de l'activation dans le modèle à travers ces trois niveaux.

Il s'agit d'un modèle à deux étapes, correspondant respectivement à la sélection lexicale et à l'encodage phonologique. Au cours de la première étape, dite de *sélection lexicale*, les traits sémantiques du mot à produire reçoivent une vague d'activation. Cette activation se propage dans tout le réseau pendant une certaine durée. Après un certain temps, l'unité mot la plus activée est sélectionnée, ce qui correspond à la première étape de sélection lexicale. La seconde étape, dite d'*encodage phonologique*, démarre avec une vague d'activation partant du mot sélectionné. Cette activation se propage vers le niveau phonologique : après un certain temps, les phonèmes les plus activés sont sélectionnés et ordonnés dans un cadre phonologique.

Ce modèle est interactif car toutes les connexions sont bidirectionnelles : l'activation se propage dans les deux sens. Ce modèle permet en particulier d'expliquer et de simuler de nombreuses erreurs de production, dont les erreurs sémantiques, formelles et mixtes. Un des arguments en faveur de l'interactivité est

l'existence *d'erreurs mixtes* (sémantiques et formelles ; Dell et Reich, 1981). Par exemple, si vous devez produire CAT mais qu'accidentellement vous produisez RAT, il s'agit d'une erreur mixte qui partage à la fois de l'information sémantique et phonologique avec la cible initiale CAT. Ce type d'erreur s'explique de la façon suivante dans le modèle de Dell que nous reproduisons figure 1.2 (Dell *et al.*, 1997). L'unité mot CAT est fortement activée par ses traits sémantiques élémentaires. Cette unité mot CAT envoie ensuite de l'activation vers les unités phonémiques correspondantes /k/, /æ/ et /t/. Certains traits sémantiques de CAT (comme « animé » et « mammifère ») coactivent l'unité mot RAT. Cette unité mot RAT reçoit également de l'activation rétroactive des phonèmes /æ/ et /t/. Cette convergence d'activations du niveau sémantique et du niveau phonologique donne à RAT plus de chance d'émerger comme erreur qu'un mot relié sémantiquement seulement (comme DOG) ou phonologiquement seulement (comme MAT). L'existence d'une interactivité dans le modèle signifie qu'au cours de la sélection lexicale, les unités mots des voisins phonologiques de CAT tels que MAT, SAT et CAN sont également activées. Selon Dell *et al.* (1997), cette activation bidirectionnelle et cette interactivité permettent une grande fluidité de la sélection lexicale des mots.

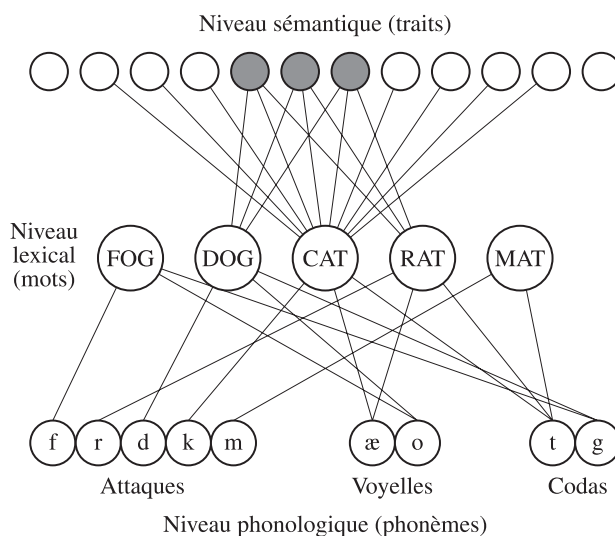


Figure 1.2. *Modèle à activation interactive en cascade de Dell et al. (1997). Le niveau sémantique est composé d'unités sémantiques correspondant à des traits sémantiques ; le niveau lexical est composé d'unités mots et le niveau phonémique est constitué de phonèmes (les attaques, les voyelles et les codas). Les connexions entre les niveaux sont excitatrices et bidirectionnelles (il n'existe pas de connexions inhibitrices). L'activation se propage en cascade de façon interactive, du niveau sémantique vers le niveau phonologique (en passant par le niveau lexical), mais aussi du niveau phonologique vers le niveau sémantique (en passant par le niveau lexical). Les traits sémantiques en commun pour CAT (chat), DOG (chien) et RAT (rat) apparaissent en noir.*

Dell (1986, 1988) a donné une description très précise des mécanismes impliqués dans l'étape d'encodage phonologique (voir cette description dans la figure 1.3). Dell (1986) suggère que les mots sont représentés dans le réseau par des unités correspondant aux morphèmes, syllabes, rimes, segments, groupements de segments et traits. La figure 1.3 illustre cela pour le mot RADAR (par simplicité, les rimes et les traits ne sont pas représentés). Par exemple, lorsque nous devons produire le mot RADAR, le morphème correspondant et ses unités segmentales /r/, /a/, /d/, /a/ et /r/ sont récupérés en mémoire et sont utilisés pour construire un programme moteur approprié. Ensuite, les segments forment l'attaque et le nucleus de la première syllabe (RA), et l'attaque, le nucleus et la coda de la seconde syllabe (DAR), respectivement. Les unités correspondant aux segments et aux groupements de segments sont donc marquées en fonction de leur position syllabique. Les unités sont connectées entre elles par des connexions bidirectionnelles de différentes forces. Au cours de l'encodage phonologique, l'activation de l'unité morphémique RADAR augmente grâce à l'activation provenant du lemma RADAR. L'activation se propage ensuite à travers le réseau, chaque unité envoyant de l'activation à ses voisins. La mise en correspondance du morphème RADAR avec ses segments phonologiques se fait par la sélection des segments ou groupements de segments les plus activés. Les unités sélectionnées sont ensuite insérées dans des cadres syllabiques. Les unités segmentales pour une syllabe sont sélectionnées en parallèle.

Dans le cas de morphèmes polysyllabiques (comme REPEINDRE), les connexions entre l'unité morphémique et les unités syllabiques sont étiquetées par ordre et position. Les syllabes sont codées de façon sérielle. Le codage successif des syllabes est accompli en augmentant temporairement le taux d'activation d'une syllabe et en diminuant celui des autres syllabes. Par la suite, Dell (1988) a ajouté un niveau supplémentaire pouvant coder toutes les formes possibles (voir figure 1.3). Chaque mot dans le réseau lexical est connecté à la tête de la forme du mot qui représente sa structure CV. L'unité correspondant à la forme du mot active de façon sérielle les unités segmentales correspondantes : l'attaque-consonne (C_{At}), la voyelle (V) et la coda-consonne (C_{Co}). La sélection des unités correspondant à l'attaque, au nucleus et à la coda se fait de manière sérielle. Cela permet au modèle de rendre compte des effets sériels de l'encodage phonologique (voir Segui et Ferrand, 2000 pour un résumé). Toutefois, Roelofs (1997b) a suggéré une autre façon de modéliser l'étape d'encodage phonologique (voir le paragraphe suivant).

Une des limitations du modèle de Dell est qu'il n'a pas été développé pour simuler les données de la chronométrie mentale mais seulement pour simuler les distributions des erreurs de production. Le modèle sériel de Levelt *et al.* (1999) permet au contraire d'expliquer certaines données de la chronométrie mentale.

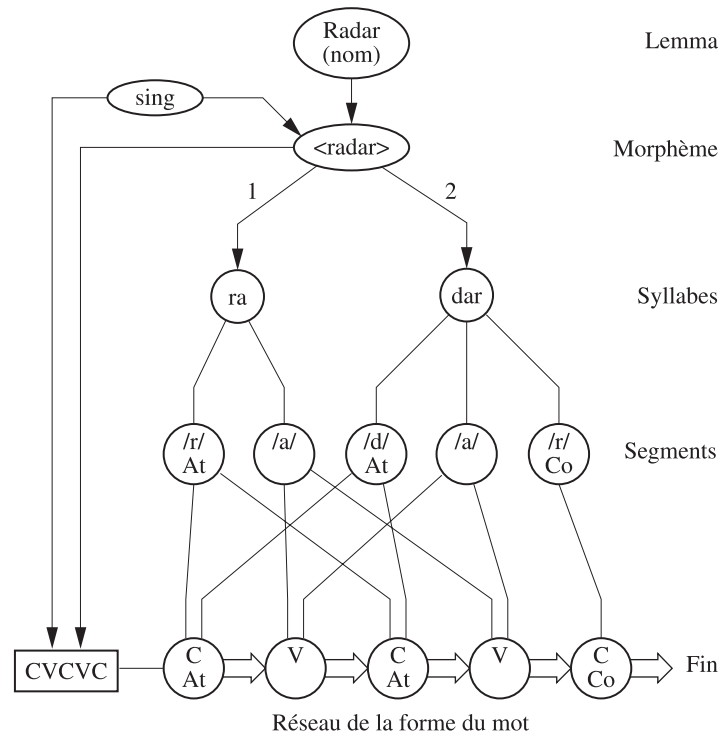


Figure 1.3. Fragment du modèle d'encodage phonologique (adapté de Dell, 1986, 1988). Par souci de simplicité, les unités correspondant aux rimes et aux traits phonétiques ne sont pas représentées

1.4.2. Le modèle strictement sériel de Levelt

Levelt et Roelofs (Levelt *et al.*, 1991, 1999 ; Roelofs, 1997b) ont proposé un modèle de la production de la parole à étapes discrètes et strictement sérielles, sans rétroaction d'activation. Dans la dernière version de ce modèle (appelé WEAVER), la production des mots passe par une série d'étapes de traitement dont chacune correspond à un niveau de représentation et de traitement spécifique. La transmission de l'information d'un niveau à l'autre opère de manière strictement sérielle : le traitement d'un type particulier d'information doit être effectué avant que ne commence le traitement de l'information associée au niveau suivant. La figure 1.4 représente un fragment du modèle WEAVER. Ce modèle comprend trois niveaux, comme le modèle de Dell *et al.* (1997) : le niveau conceptuel/sémantique, le niveau syntaxique (des lemmas, avec les propriétés syntaxiques comme le genre, la catégorie grammaticale, etc.), le niveau phonologique (des lexèmes, avec la structure métrique, les morphèmes, les segments et les syllabes).

Comment fonctionne WEAVER ? Supposons que l'on doive dénommer l'image d'un MOUTON. La première étape consiste à activer le concept lexical MOUTON. L'activation se propage alors du niveau conceptuel/sémantique au niveau des lemmas. Dans le réseau conceptuel/sémantique, l'activation se diffuse vers les concepts reliés (comme LAMA et CHEVRE). Ces concepts envoient à leur tour de l'activation à leurs lemmas. La probabilité de sélectionner le lemma cible MOUTON est fonction du niveau d'activation du lemma MOUTON par rapport au niveau d'activation total de tous les autres lemmas activés (comme LAMA et CHEVRE). Autrement dit, il existe une compétition entre les lemmas sémantiquement reliés. Ce mécanisme de compétition permet d'expliquer un certain nombre de résultats comme les erreurs sémantiques et les effets d'inhibition sémantique (Segui et Ferrand, 2000). Un seul lemma est sélectionné (MOUTON ici). La sélection de ce lemma permet ensuite la récupération des codes morphophonologiques (comme la structure métrique et segmentale ; voir Segui et Ferrand, 2000 pour plus de détails). La récupération des informations morphophonologiques succède donc à la sélection du lemma. Lorsque le mot cible est CAT, seul le lemma CAT se trouve sélectionné et envoie de l'activation à son code phonologique /kaet/. Les autres lemmas (comme DOG et RAT) ne sont pas sélectionnés et n'envoient pas d'activation au niveau phonologique.

Le niveau phonologique est constitué de trois types d'unités différentes (voir figure 1.4) : les unités morphémiques (<select>), les unités segmentales (/s/, /i/, /l/, /ε/, /k/ et /t/), et les unités syllabiques ([si], [lɛk], et [lɛkt]). Les unités morphémiques permettent de coder les racines et les affixes. Ces unités morphémiques sont connectées au lemma et à ses paramètres diacritiques. La racine morphologique <select> est connectée aux segments phonologiques la définissant, ainsi qu'à sa structure métrique (nombre de syllabes et accentuation). Point important, l'information métrique n'indique pas quels segments constituent telle ou telle syllabe, de même qu'elle ne spécifie pas la forme CV (contrairement au modèle de Dell). Les liens entre le morphème et ses unités segmentales indiquent la position sérielle des segments dans le morphème (dans l'exemple choisi ici, le /s/ est en première position, le /i/ en deuxième position, etc.). Les positions possibles des segments dans les syllabes (attaque, nucleus, coda) sont précisées par les liens entre les unités segmentales et les unités syllabiques. Par exemple, les liens dans le réseau spécifient que /l/ est l'attaque de la syllabe [lɛk] et de la syllabe [lɛkt]. L'information dans le réseau est récupérée via la propagation de l'activation. L'encodage phonologique démarre dès que l'unité morphémique reçoit l'activation de son lemma. Celle-ci se propage ensuite de façon unidirectionnelle à travers le réseau. Chaque unité envoie une partie de son activation à ses voisins immédiats. Il existe également un déclin spontané de l'activation. Ce modèle diffère de celui de Dell en ce que la syllabation y est calculée par des règles tandis qu'elle est stockée directement dans le lexique mental dans le modèle de Dell.

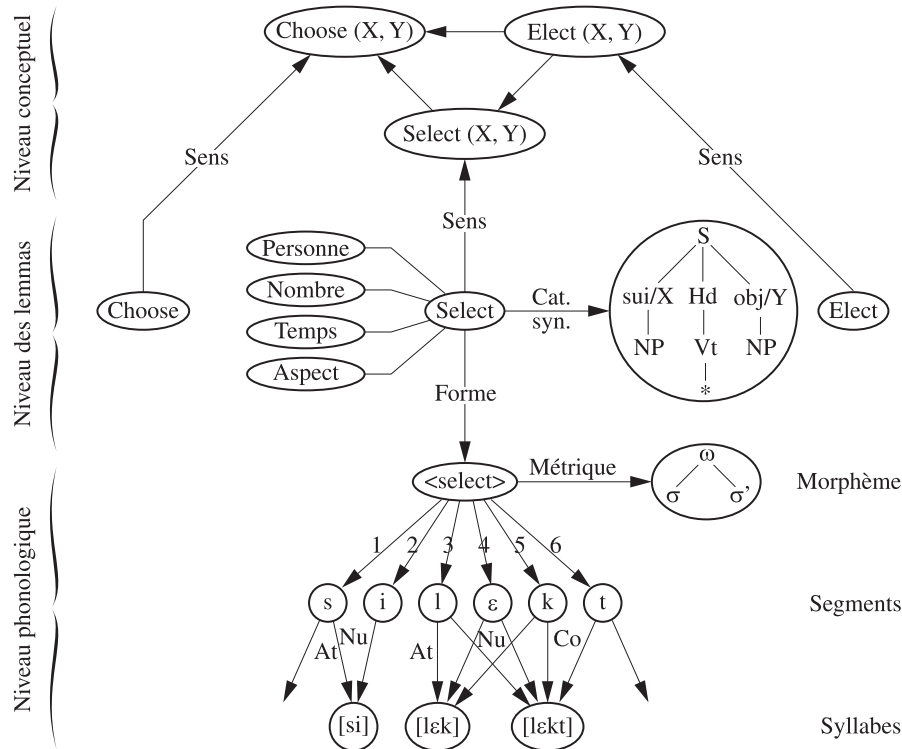


Figure 1.4. WEAVER : modèle sériel de Levelt et al. (1999). La production de la parole passe par plusieurs étapes sérielles : l'étape conceptuelle, l'étape de sélection lexicale (correspondant à la récupération du lemme) et l'étape d'encodage phonologique. Les nœuds du niveau conceptuel représentent les concepts lexicaux et les connexions entre ces concepts représentent les relations sémantiques. Au niveau des lemmas (le niveau syntaxique), chaque nœud représente un lemme, c'est-à-dire les propriétés syntaxiques des mots (comme le nombre, le genre, etc.). La connexion du lemme avec son nœud conceptuel représente le sens du mot. Le lemme est également connecté à sa forme phonologique (ou lexème). L'activation se propage de façon unidirectionnelle des lemmas vers le niveau phonologique. Seul le lemme sélectionné peut envoyer de l'activation au niveau phonologique.

Ce modèle est extrêmement détaillé et rend compte d'un grand nombre de résultats. Toutefois, comme nous allons le voir, sa nature strictement sérielle a été remise en question par de nombreux auteurs.

1.4.3. Sérialité ou interactivité ?

La conception strictement sérielle du modèle de Levelt *et al.* (1991, 1999) a été remise en question par de nombreux travaux expérimentaux (Cutting et Ferreira, 1999 ; Dell *et al.*, 1997 ; Jescheniak et Schriefers, 1998 ; Peterson et Savoy, 1998 ;

Rapp et Goldrick, 2000 ; voir Segui et Ferrand, 2000, pour une synthèse). En particulier, Rapp et Goldrick (2000) ont testé quatre modèles de la production de la parole à l'aide de simulations (voir figure 1.5).

Le *modèle standard discret* (figure 1.5a) se caractérise principalement par une activation strictement propagée vers le niveau supérieur, sans rétroaction (Levelt et al., 1999), un traitement limité à l'étape considérée et, enfin, un processus de sélection à la fin de chaque étape. De plus, seul l'item sélectionné à la fin d'une étape donnée est traité par l'étape suivante. Autrement dit, le modèle standard n'a ni activation en cascade, ni rétroaction d'un niveau vers un autre. Par exemple, l'étape 1 (sélection lexicale) commence lorsque l'information sémantique concernant le mot cible (CAT) active la cible et ses compétiteurs sémantiquement reliés (comme RAT et DOG) au niveau lexical. Cette étape se termine lorsqu'une seule unité lexicale est sélectionnée (ici CAT) : l'activation des unités lexicales en compétition (comme DOG et RAT) se voit réduite à zéro. Au cours de l'étape 2 (encodage phonologique), la récupération des phonèmes a lieu uniquement pour l'unité lexicale sélectionnée (CAT ; figure 1.5a).

Le *modèle en cascade sans interaction* (figure 1.5b). Comme pour le modèle standard, l'activation est propagée vers l'avant sans rétroaction. Toutefois, l'activation n'est pas limitée à l'étape considérée mais est transmise en cascade aux niveaux suivants. Enfin, bien qu'il existe des moments de sélection à la fin de chaque étape, le traitement de tous les items se poursuit au cours des étapes suivantes. Par exemple, l'étape 1 (sélection lexicale) commence lorsque l'information sémantique concernant le mot cible (CAT) active la cible et ses compétiteurs sémantiquement reliés (comme RAT et DOG) au niveau lexical. Plusieurs unités lexicales sont sélectionnées, dont le mot cible (CAT) ainsi que ses compétiteurs lexicaux (DOG et RAT). Au cours de l'étape 2 (encodage phonologique), l'activation des phonèmes concerne à la fois l'unité lexicale cible (CAT) et les compétiteurs (DOG et RAT ; figure 1.5b).

Le *modèle interactif restreint* (figure 1.5c) : l'activation est transmise en cascade vers les autres niveaux, comme pour le modèle en cascade, mais il existe une rétroaction de l'activation du niveau phonologique vers le niveau sémantique (comme dans le modèle de Dell). Par exemple, l'étape 1 démarre lorsque l'information sémantique concernant la cible CAT active l'unité lexicale CAT et ses compétiteurs lexicaux DOG et RAT. De plus, l'étape 1 continue pendant que toutes les unités lexicales concernées (CAT, DOG et RAT) envoient de l'activation aux phonèmes correspondants. Les étapes 1 et 2 impliquent non seulement une propagation de l'activation vers l'avant, mais aussi une rétroaction de l'activation du niveau phonologique vers le niveau lexical. L'étape 1 se termine avec la sélection du candidat le plus activé au niveau lexical (CAT dans notre exemple), mais l'activation des compétiteurs lexicaux (DOG et RAT) n'est pas remise à zéro. Pendant l'étape 2, le traitement de tous les niveaux se poursuit jusqu'à la fin de cette étape, au cours de laquelle les phonèmes les plus activés vont être sélectionnés.

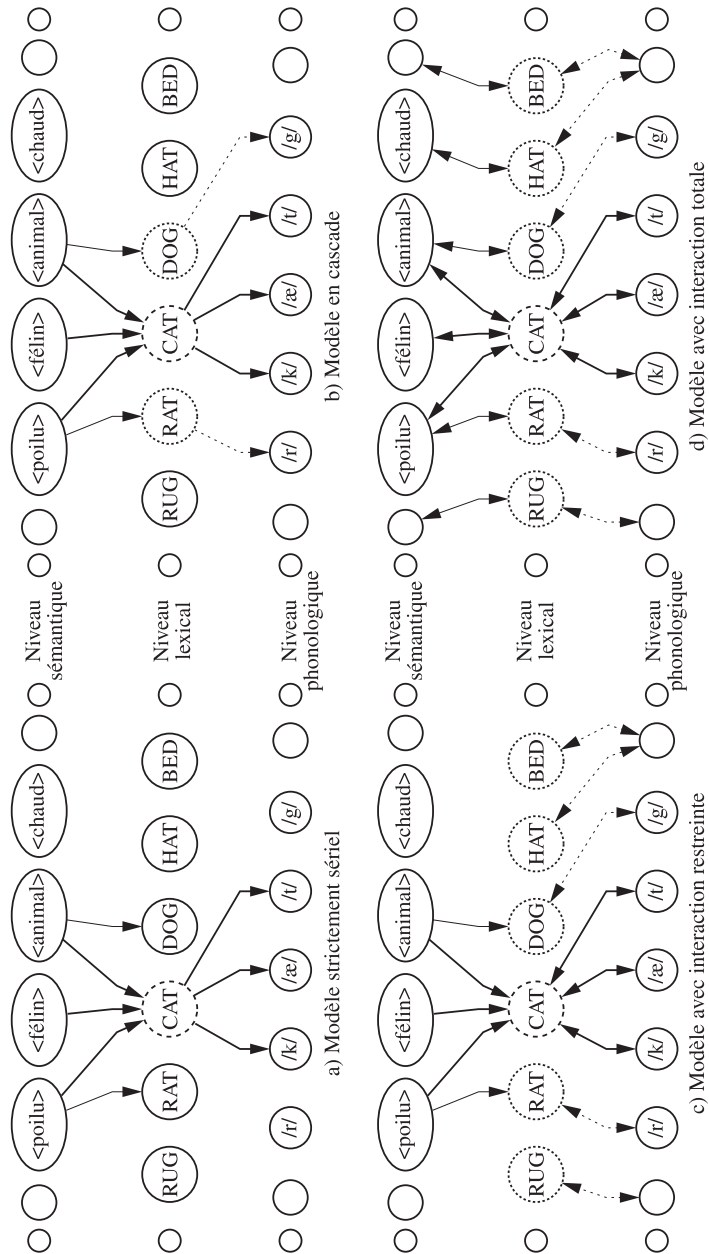


Figure 1.5. Représentations schématiques de quatre modèles possibles de la production de la parole pour le mot CAT (d'après Rapp et Goldrick, 2000). Les flèches épaisses correspondent à un flux d'activation impliquant le mot cible CAT, les flèches fines correspondent à un flux d'activation impliquant les voisins lexicaux, et les flèches en pointillé correspondent à un flux d'activation impliquant des mots différents du mot cible.

Le *modèle interactif exhaustif* (figure 1.5d). Il possède les mêmes caractéristiques que le modèle précédent, sauf que l'interaction existe à tous les niveaux, entre les niveaux phonologique et lexical, et entre les niveaux lexical et sémantique.

Les simulations de Rapp et Goldrick (2000) ont porté sur quatre faits : les erreurs mixtes observées chez les sujets normaux ; l'effet de biais lexical ; les erreurs sémantiques ; les erreurs phonologiques. Pour pouvoir simuler les résultats, Rapp et Goldrick montrent qu'un modèle doit posséder au minimum : un mécanisme d'activation en cascade et un mécanisme interactif entre le niveau phonologique et le niveau sémantique. Les résultats de ces simulations vont donc dans le sens d'un modèle en cascade avec une interactivité limitée aux niveaux sémantique et phonologique (figure 1.5c).

Que choisir entre un modèle strictement sériel et un modèle interactif ? Selon Levelt *et al.*, « la sélection lexicale et l'encodage phonologique remplissent deux fonctions très différentes. La première vise à une recherche rapide du mot approprié dans un lexique énorme ; la seconde vise à créer un programme articulaire pour l'item sélectionné. Toute rétroaction du niveau phonologique vers le niveau sémantique fragiliserait le système et conduirait à un nombre d'erreurs bien supérieur à celui qui est attesté dans une communication normale. La modularité, peut-on dire, est une protection naturelle contre cette dérive » (Levelt *et al.*, 1991, p. 618). Au contraire, selon Dell *et al.* (1997), l'interactivité rend le système plus robuste et moins sujet à l'erreur. En effet, il serait plus avantageux pour le locuteur de choisir un lemme dont la forme phonologique sera plus facile à récupérer par la suite. Les lemmes dont les formes sont accessibles reçoivent plus d'activation (*via* la rétroaction d'activation du niveau phonologique vers le niveau sémantique) que ceux dont les formes ne sont pas accessibles. De plus, l'interactivité permet de satisfaire de nombreuses contraintes et rend le système optimal dans la mesure où il gère ces contraintes multiples de façon satisfaisante. Les données empiriques, temps de réaction et erreurs de production, ainsi que les simulations sur ordinateur favorisent plutôt le modèle interactif restreint.

1.4.4. Le modèle à réseau indépendant de Caramazza

Comme nous l'avons souligné auparavant, la distinction entre le niveau sémantique/syntaxique (le niveau des lemmes) et le niveau phonologique (le niveau des lexèmes) est largement acceptée. Toutefois, Caramazza et Miozzo (Caramazza, 1997 ; Caramazza et Miozzo, 1997) proposent un modèle différent reposant sur les données de la Neuropsychologie et du phénomène du mot sur le bout de la langue. De récents travaux ont montré qu'il était possible de récupérer l'information syntaxique d'un mot (comme le genre grammatical par exemple) dans des cas où il était impossible de récupérer la moindre information phonologique relative à ce

même mot : ces résultats ont été obtenus avec des sujets normaux en état de mot sur le bout de la langue (Ferrand, 2001c ; Vigliocco *et al.*, 1997), ainsi qu'avec des patients aphasiques (Badecker, Miozzo et Zanuttini, 1995). Ce résultat est compatible avec les modèles à deux étapes. Toutefois, Caramazza et Miozzo (1997) ont montré que la récupération des propriétés syntaxiques n'est pas corrélée avec l'habileté à récupérer l'information phonologique partielle des mots. En particulier, dans une série d'expériences utilisant le phénomène du mot sur le bout de la langue, Caramazza et Miozzo ont montré que l'information phonologique partielle relative à un mot sur le bout de la langue pouvait être récupérée même lorsque le locuteur était incapable de récupérer la moindre information syntaxique concernant ce mot. Le modèle de Levelt *et al.* (1999) prédit qu'il ne devrait pas être possible de récupérer l'information phonologique partielle sans récupérer auparavant l'information syntaxique. Or, Caramazza et Miozzo (1997) montrent que tel est le cas.

Caramazza (1997) a proposé un modèle de la production de la parole dans lequel l'information syntaxique est *indépendante* de l'information sémantique et de l'information phonologique. Ce modèle (voir la figure 1.6) suggère que les connaissances lexicales sont organisées en réseaux indépendants mais interconnectés. Le réseau lexical-sémantique représente le sens des mots comme des ensembles de propriétés sémantiques, de traits sémantiques ou de prédicats. Le réseau syntaxique représente les traits syntaxiques des mots comme la catégorie grammaticale, le genre, le type d'auxiliaire, le temps, etc. Dans ce réseau, les nœuds sont organisés en sous-réseaux correspondant à différentes fonctions syntaxiques. Par exemple, un sous-réseau correspond à des catégories (nom, verbe, adjectif, etc.), un autre au genre (masculin, féminin), un autre encore au type d'auxiliaire (avoir, être), etc. Les nœuds à l'intérieur du sous-réseau syntaxique ont des connexions inhibitrices car ils sont en compétition. Le réseau des lexèmes représente la forme phonologique des items lexicaux.

Ce modèle à réseaux indépendants partage de nombreuses propriétés avec d'autres modèles de l'accès au lexique : l'information lexicale-sémantique est représentée indépendamment des représentations syntaxiques et phonologiques (comme dans le modèle de Dell) ; le réseau lexical-sémantique est componentiel (comme dans Dell, mais contrairement à Levelt) ; il s'agit d'un modèle dont l'activation se propage uniquement vers l'avant, sans rétroaction (comme dans Levelt, et contrairement à Dell ; Rapp et Goldrick, 2000). Toutefois, contrairement aux modèles précédents, l'activation se propage *simultanément* et *indépendamment* du réseau lexical-sémantique vers le réseau syntaxique d'une part, et vers le réseau des lexèmes d'autre part. Autrement dit, ce modèle n'adhère pas à l'hypothèse de *la médiation syntaxique obligatoire* de Levelt *et al.* (1999). Les données issues de la neuropsychologie cognitive suggèrent effectivement que l'information lexicale-sémantique et l'information grammaticale sont indépendantes puisque une lésion peut affecter la première et pas la seconde, ou au contraire, la seconde et pas la première.

L'information syntaxique et l'information phonologique sont indépendantes pour les mêmes raisons. Les hypothèses centrales du modèle de Caramazza (1997) sont les suivantes :

- les représentations lexicales-sémantiques, syntaxiques et phonologiques sont stockées indépendamment dans des réseaux séparés ;
- les représentations phonologiques (les lexèmes) sont activées indépendamment par les représentations sémantiques ;
- les représentations lexicales-sémantiques sont componentielles et activent en parallèle l'information syntaxique et l'information phonologique ;
- la représentation lexicale-sémantique active en parallèle tous les lexèmes des mots partageant des traits sémantiques avec le lemma sélectionné.

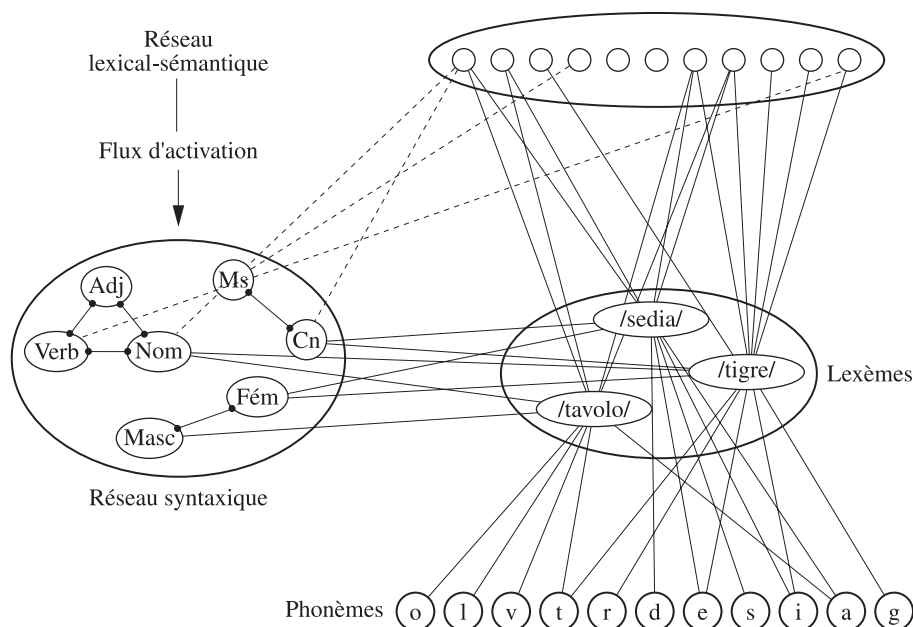


Figure 1.6. Représentation schématique du réseau indépendant (d'après Caramazza, 1997) montrant les relations entre les représentations lexicales-sémantiques, syntaxiques et phonologiques (lexèmes). Le flux d'activation se propage du réseau lexical-sémantique au réseau syntaxique et aux lexèmes, puis aux phonèmes. Les lignes en pointillé indiquent une faible activation. Les connexions à l'intérieur d'un réseau sont inhibitrices.

Caramazza suggère que la sélection des représentations (lexicales) sémantiques ne garantit pas l'accès aux informations syntaxiques et que l'accès aux informations phonologiques d'un mot ne dépend pas strictement de l'accès préalable à ses

informations grammaticales. Contrairement aux modèles de Dell *et al.* (1997) et de Levelt *et al.* (1999), le modèle de Caramazza reste « verbal » pour le moment et n'a pas été implémenté sur ordinateur. Il reste donc difficile à comparer directement avec les deux modèles précédents et nous ne pouvons faire que des prédictions verbales peu précises. De plus, le modèle à réseau indépendant ne fournit aucun détail concernant les mécanismes impliqués au niveau de l'encodage phonologique.

1.5. Conclusion

Dans ce chapitre, nous avons abordé brièvement les trois modèles principaux de la production de la parole : le modèle en cascade interactif de Dell, le modèle strictement sériel de Levelt et Roelofs, et le modèle à réseau indépendant de Caramazza et Miozzo. Nous avons vu qu'il existait une controverse importante concernant la nature sérielle ou interactive des relations entre les niveaux sémantique et phonologique. La plupart des travaux actuels favorisent plutôt les conceptions interactives.

Nous n'avons pas abordé la neuropsychologie de la production de la parole mais c'est un domaine qui se développe extrêmement rapidement et qui apporte d'ores et déjà des contraintes importantes sur la modélisation des mécanismes impliqués dans la production de la parole (Dell *et al.*, 1997; Rapp et Goldrick, 2000). Nous n'avons pas non plus abordé les modèles de la lecture à voix haute, bien qu'à l'évidence ces deux tâches (lecture à voix haute et production de la parole) partagent de nombreux processus. Peut-on pour autant dire qu'un modèle de la production de la parole est un modèle de la lecture à voix haute, et *vice versa* ? Enfin, certaines questions importantes n'ont toujours pas trouvé de réponse. En particulier, comment le système de production de la parole est-il lié au système de perception de la parole (voir Segui et Ferrand, 2000 pour un début de réponse) ? Quelles sont les régions cérébrales impliquées dans les processus de préparation conceptuelle, de sélection lexicale, d'encodage phonologique, etc. ? A l'avenir, les modèles de la production de la parole vont-ils tendre vers des modèles à traitement parallèle distribué (Dell *et al.*, 1993) et remplacer les modèles symboliques classiques ?

1.6. Bibliographie

- Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, 57, 193-216.
- Berg, T., & Schade, U. (1992). The role of inhibition in a spreading-activation model of language production: I. The psycholinguistic perspective. *Journal of Psycholinguistic Research*, 21, 405-434.

- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177-208.
- Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the “tip-of-the-tongue” phenomenon. *Cognition*, *64*, 309-343.
- Cutting, J.C., & Ferreira, V.S. (1999). Semantic and phonological flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 318-344.
- Dell, G.S., & Reich, P.A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*, 611-629.
- Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283-321.
- Dell, G.S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124-142.
- Dell, G.S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149-195.
- Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., & Gagnon, D.A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801-838.
- Ferrand, L. (1994). Accès au lexique et production de la parole : Un survol. *L'Année Psychologique*, *94*, 295-312.
- Ferrand, L. (1997). La dénomination d'objets : Théories et données. *L'Année Psychologique*, *97*, 113-146.
- Ferrand, L. (2001a). *La production du langage*. Grenoble : Presses Universitaires de Grenoble.
- Ferrand, L. (2001b). La production du langage: une vue d'ensemble. *Psychologie Française*, *46*, 3-15.
- Ferrand, L. (2001c). Grammatical gender is also on the tip of French tongues. *Current Psychology Letters*, *5*.
- Fromkin, V.A. (1973). *Speech errors as linguistic evidence*. La Haye : Mouton.
- Jescheniak, J.D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1256-1274.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA : MIT Press.
- Levelt, W.J.M., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122-142.
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.

- Peterson, R.R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 539-557.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107, 460-499.
- Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, 35, 854-876.
- Roelofs, A. (1997a). A case for non-decomposition in conceptually driven word retrieval. *Journal of Psycholinguistic Research*, 26, 33-67.
- Roelofs, A. (1997b). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249-284.
- Rossi, M., & Peter-Defare, E. (1998). *Les lapsus, ou comment notre fourche a langué*. Paris : Presses Universitaires de France.
- Schriefers, H., Meyer, A.S., & Levelt, W.J.M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, 29, 86-102.
- Segui J., & Ferrand L. (2000). *Leçons de parole*. Paris : Odile Jacob.
- Stemberger, J.P. (1985). An interactive activation model of language production. In W. Ellis (Ed.), *Progress in the psychology of language, vol. 1*. Hillsdale : Erlbaum.
- Vigliocco, G., Antonini, T., & Garrett, M.F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8, 314-317.