

# Chapitre 1

## De la linguistique à la psycholinguistique. . .

*“Le fait est que, de nature et originellement, aucun nom n’appartient à rien en particulier, mais bien en vertu d’un décret et d’une habitude, à la fois de ceux qui ont pris cette habitude et de ceux qui ont décidé l’appellation”*

Platon, Le Cratyle.

### 1.1 Introduction

Le problème que toutes les langues naturelles ont à résoudre est de faire correspondre du sens à des formes. Trois raisons principales rendent la résolution de la correspondance formes-sens difficile : i) le domaine du sémantique est vaste et complexe, ii) la variabilité des formes possibles est très grande, iii) la mise en correspondance du sens et des formes pour l’exprimer peut se faire selon de très nombreuses modalités de par la complexité même des espaces qu’elle relie. Le principe de la double articulation (Martinet 1960), que nous observons dans toutes les langues du monde, résout ces difficultés. La première articulation a permis, d’une part, d’exprimer des formes holistiques qui sont sémantiquement simples et, d’autre part, la mise en place, au cours de l’évolution humaine, de systèmes de

communication basés sur des conventions partagées par les locuteurs d’une langue. Dans Le cours de linguistique générale (1916), Ferdinand de Saussure fait l’observation que les langues naturelles sont en effet des systèmes “immotivés”, c’est-à-dire des systèmes dans lesquels la relation entre la forme d’un mot et sa signification est arbitraire et conventionnelle. Par exemple, la forme *sœur* en français qui exprime le concept de “sœur”, ne présente aucune particularité justifiant son usage : rien dans les sons employés n’est relié au concept biologique et/ou culturel de “sœur”. Afin de rendre compte de la nature des signes linguistiques, Ferdinand de Saussure utilise la métaphore d’une pièce à deux faces, avec d’un côté le concept - le signifié, et de l’autre l’image acoustique ou graphique qui lui est associée - le signifiant (voir Figure 1.1).

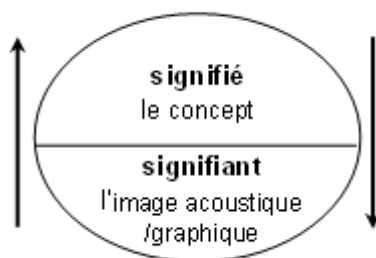


FIG. 1.1 – Description du signe linguistique selon Ferdinand de Saussure (1916). Le signe linguistique est une entité abstraite à deux faces formée par la réunion d’un signifié (le concept) et d’un signifiant (image acoustique ou graphique).

Des systèmes langagiers basés sur un système de signes arbitraires ne permettent de résoudre que partiellement le problème de la correspondance entre sens et formes. En effet, le nombre de messages que nous pouvons être amenés à communiquer est virtuellement infini et le système d’arbitrarité dès lors insuffisant : trop de signes -en fait un nombre virtuellement infini- seraient nécessaires et rendraient les langues trop complexes pour être utilisées par leurs locuteurs. La charge mnésique deviendrait simplement très vite trop importante.

Pour palier ces difficultés, la segmentation des unités holistiques pourvues de sens en

unités plus petites est un processus intéressant. En effet, à partir de cette segmentation, il est possible de former de nouvelles unités pourvues d'une nouvelle signification, mais basées sur des unités déjà existantes. Ainsi, pour exprimer des messages complexes, on ne crée pas pour chaque message une forme spécifique et holistique, mais on compose à partir d'unités plus simples : on compose en particulier les mots à partir de formes linguistiques minimales dotées d'une signification propre, les morphèmes (Bloomfield, 1933). Par exemple, le mot *chanteuse* est composé de deux morphèmes : *chant-* et *-euse*. Ces morphèmes peuvent être eux-mêmes segmentés en unités minimales dépourvues de sens (deuxième articulation), les phonèmes.

Au cours de cette thèse, nous allons plus particulièrement nous intéresser à la deuxième articulation et à la composition des mots polymorphémiques. Les langues qui sont concernées par le polymorphisme sont appelées soit **langues fusionnelles**, dont les langues indo-européennes sont très représentatives (par ex. l'allemand, l'anglais, l'espagnol, le français...), soit **langues agglutinantes** (par ex. le finnois, le turc...). La distinction entre ces deux types de langues est la suivante : dans les langues fusionnelles, la combinaison de formes linguistiques permet d'ajouter plusieurs propriétés à une base morphémique de manière simultanée ; dans le cas de *chant-* + *-euse*, les informations "faire de son métier" et "nom féminin" sont simultanément ajoutées à la base *chant* grâce au suffixe *-euse*. Dans les langues dites agglutinantes, la création de mots polymorphémiques se réalise en "collant" les morphèmes les uns derrière les autres, de telle sorte que les frontières entre eux demeurent bien nettes et que chaque morphème corresponde à un seul trait sémantique ou fonctionnel. Pour illustrer ceci, nous pouvons prendre le cas du finnois ou du turc, deux langues qui sont très représentatives de ce mécanisme. Par exemple, en finnois, la forme *taloissani* ('dans mes maisons') peut se segmenter de la façon suivante : *talo*, 'maison' ; *i*, la marque du pluriel ; *ssa*, 'dans' ; *ni*, suffixe qui marque la possession à la première personne (*mon/ma* en français). En turc, la construction est très similaire : à partir de *ev* ('maison'), on peut former les mots *evler* ('les maisons'), *evlerim* ('mes maisons'), et *evlerimde* ('dans mes maisons').

Au cours de cette thèse, nous nous sommes intéressés de manière privilégiée à la composition des mots polymorphémiques au sein des langues fusionnelles, et plus particulièrement en français. Nous allons voir maintenant que deux approches permettent d'étudier la structure des mots polymorphémiques : l'approche syntagmatique et l'approche paradigmatique.

### L'approche syntagmatique

L'approche syntagmatique permet d'étudier les mots polymorphémiques selon un axe "horizontal" : les morphèmes sont considérés comme des "briques de construction" posées les unes après les autres. Plusieurs types de "briques" ou morphèmes peuvent se différencier. On distingue généralement les **racines** des **affixes**.

### Les racines

Si nous nous intéressons tout d'abord aux racines, nous faisons l'observation qu'elles sont "*l'élément irréductible récurrent dans les formes lexicales apparentées par le sens et considéré en linguistique historique comme la forme la plus ancienne expliquant tous les dérivés ultérieurs*" (TLFI, Trésor de la langue française informatisé, <http://atilf.atilf.fr/tlf.htm>). Par exemple, dans le mot *campeur*, la suite de lettres *camp* correspond à la racine. Ce terme "racine" peut être admis comme terme générique pour rendre compte de la base à partir de laquelle sont construits les mots polymorphémiques. Cependant, nous pouvons faire les distinctions suivantes :

1. Dans le cas où la suite de phonèmes correspond à un morphème qui peut apparaître de manière isolée dans le discours, on parle de **morphème libre** ou **lexical**. Dans l'exemple *campeur*, *camp-* est une racine libre.
2. Les racines (mais par forcément toutes) peuvent être accompagnées d'un allongement pour constituer un "thème" (Booij, 2005), et celui-ci peut être le point d'an-

crage pour la construction de nouveaux mots. Par exemple, dans le mot *format*, l’allongement *-at* est adjoint à la partie initiale *form-*, et leur assemblage constitue un thème sur lequel sont construits les mots *format-ion*, *format-eur*, *in-format-if*.

3. Quand la suite de phonèmes est un morphème qui ne peut apparaître de manière isolée dans un discours, on parle soit de **morphème lié** soit de **radical** pour être plus précis<sup>1</sup>. Dans le mot *navigation*, *navig-* est un radical.

## Les affixes

Dans les mots polymorphémiques, les racines sont accompagnées d’un ou plusieurs affixe(s). L’affixe est “l’élément lexical qui s’ajoute à un mot ou à un radical pour en modifier le sens ou la fonction”. Il s’agit toujours d’un morphème lié (il ne se produit jamais de manière isolée dans le discours). L’affixe est un **suffixe** s’il se situe après la racine (*-eur* dans *chanteur* est un suffixe), ou un **préfixe** s’il se situe avant (*dé-* dans *déchanter* est un préfixe). Quand un préfixe et un suffixe apparaissent toujours ensemble on désigne cette paire par le terme de **circomfix**. Par exemple, en allemand, le mot *ge-sing-e* (‘chantant’ en français) contient le circomfix (*ge - e*). Dans d’autres langues comme le Khmu du Laos, l’affixe n’est pas inséré de manière linéaire et on parle alors d’**infixe** : par exemple, à partir du mot *ska :t* (‘rude’ en français), on construit le mot *rudesse* en insérant l’infixe *-m-* comme suit : *s-m-ka :t*.

Parmi les suffixes, certains apportent des indications de type grammatical (genre, nombre, mode, temps, personne, cas, etc. . .) à la racine à laquelle ils sont adjoints ; ces suffixes sont nommés suffixes flexionnels (par opposition aux suffixes dérivationnels). Pour illustrer ce qu’est la **morphologie flexionnelle** nous pouvons prendre l’exemple des verbes conjugués. Ces formes verbales ne correspondent pas à des mots nouveaux à proprement parler, car généralement très transparentes au niveau de leur sens ; le suffixe de ces formes apporte uniquement des indications de temps, de mode, d’aspect ou encore de personne à la racine. Dans le mot *mangeait*, le suffixe *-ait* indique la conjugaison du verbe *manger* à

---

<sup>1</sup>Nous allons voir qu’un morphème lié peut aussi convenir pour parler des affixes.

la troisième personne du singulier du temps imparfait et du mode indicatif.

## L'allomorphie

Certaines racines ou affixes peuvent connaître des transformations phonologiques qui engendrent des allomorphes. **L'allomorphie** est le phénomène selon lequel, à partir de règles phonologiques, un morphème présente au moins deux formes de réalisations phonologiques et/ou orthographiques différentes. Par exemple en anglais, *san-* se prononce [sem] dans *sanely* [seməli] et [saen] dans *sanity* [saenəti]. *san-* possède ainsi deux réalisations phonologiques distinctes, [sem] et [saen], selon la règle appelée en anglais le “*trisyllabic shortening*” ou “*trisyllabic laxing*”. Cette règle énonce que plus il y a de syllabes dans un mot (par ex. plus il y a de suffixes placés à la suite d'un radical), plus la durée de la voyelle du radical est réduite (voir par ex. Katamba, 1989). En français, nous pouvons de même observer les cas de *frère/frérot* ou *aigle/aiglon*.

Un autre cas dans lequel la racine subit un changement de forme, sans que celui-ci puisse être explicité par une règle phonologique ou morpho-phonologique, est celui de la **suppléance**. Par exemple, en anglais, dans la paire de mots *good-better*, on a un radical *bet*<sup>2</sup> qui est d'une part phonologiquement très différent de la racine *good*, et qui d'autre part ne peut être expliqué par une quelconque règle morpho-phonologique. En français, nous avons le cas des mots *fleur/floraison/flétrissement*, de *aveugle/cécité*, ou encore de *cheval/équestre*. Les formes dans ces couples de mots sont vraiment différentes car les mots polymorphémiques sont construits sur la base de radicaux distincts (d'origines diverses) mais néanmoins synonymes du mot auquel ils sont apparentés. Par exemple, le mot *cheval* vient du latin *caballus*<sup>3</sup> qui fut d'abord la désignation populaire péjorative qui s'appliquait à un mauvais cheval, et qui s'est ensuite généralisé, remplaçant le mot *equus* du latin classique. *Caballus* et *equus* sont ainsi devenus synonymes (Dictionnaire le Robert,

---

<sup>2</sup>En anglais *-er* est un suffixe superlatif, par exemple le mot anglais *short-shorter* ‘court-plus court’.

<sup>3</sup>Avec des origines étymologiques hongroise ou celtique.

Dictionnaire Historique de la Langue Française, vol 1, p. 729-730). Ensuite, sur la base du mot *cheval*, nous retrouvons les mots *chevalin* et *chevalier*, et sur la base d'*equus* les mots *équestre* et *équitation*<sup>4</sup>. Ces derniers cas correspondent à des constructions dites “savantes” par opposition à celles dites “populaires”, c’est-à-dire construites sur la base d’une racine latine. Ce type de construction se retrouve aussi en anglais, avec par exemple les mots *horsy* (‘chevalin’ en français) et *equitation* (‘équitation’); cependant, en comparaison du français, l’anglais est une langue dans laquelle beaucoup moins de mots polymorphémiques sont construits à partir de racines d’origines différentes. C’est ainsi qu’en français on a l’expression “un témoin oculaire”, avec le mot *oculaire* construit sur la base latine *oculus*, alors que l’expression équivalente en anglais est littéralement ‘témoin d’œil’ : “*eye witness*”. Un autre cas est celui de “gaz lacrymogène”, avec le mot *lacrymogène* qui dérive de la racine latine *lacrima/lacrymal* (‘larme’) et qui correspond à “*tear gas*” en anglais, littéralement ‘gaz à larmes’.

Un dernier cas concernant la forme des racines est celui pour lequel leur sens n’est pas identifiable. Par exemple, en anglais, il existe de nombreux cas dans lesquels les mots sont considérés comme étant polymorphémiques : *conceive, deceive, perceive, receive*...ou *admit, permit, remit, transmit*...ou encore *adduce, deduce, induce, reduce, produce*... Des éléments comme *ad- per-, re-*...sont des préfixes habituels de la langue, et suggèrent donc une décomposition possible des mots précédents. Cependant, des racines comme *-ceive, -duce* ou *-mit* n’ont pas de sens clairement identifiable.

En ce qui concerne les affixes, il est également possible d’observer des cas d’allomorphie. Par exemple, les préfixes *al-* et *ap-* en italien sont deux allomorphes du préfixe *ad-* dont la consonne finale /d/ est assimilée à la première consonne de la racine qui suit : par exemple *larg-o* (‘large’) devient *al-larg-a-re* (‘élargir’) et *profond-o* (‘profond’) devient *ap-profond-i-re* (‘approfondir’). En français, il existe des cas d’allomorphie pour les suffixes lorsqu’ils sont inaccentués (le suffixe est lui-même suivi d’un autre suffixe) : *dangereux/dangerosité*; *temporel/temporalité*; *africain/africaniste*.

---

<sup>4</sup>Nous pouvons aussi ajouter la base grecque *hippos* ‘cheval’, que l’on retrouve dans le mot *hippodrome*.

## L'approche paradigmatique

Selon l'approche paradigmatique, les variations sémantiques des mots polymorphémiques sont étudiées en rapport avec la modification formelle et sémantique apportée par les affixes au morphème de base (voir Figure 1.2). Le suffixe *-eur* par exemple a pour fonction d'ajouter le sens de "quelqu'un qui pratique" au nom auquel il se rattache : *chanteur*, *inventeur*, *danseur* ect. alors que le suffixe *-asse* ajoute une valeur d'amplification : *paperasse*, *fadasse*, *mollasse*.

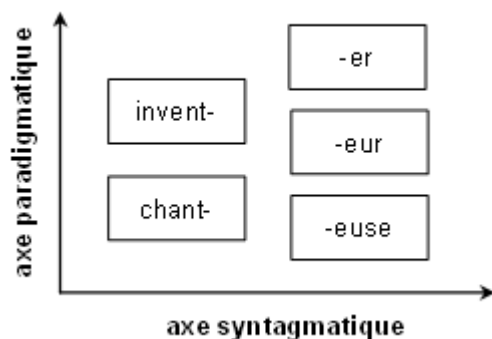


FIG. 1.2 – Schéma représentant l'axe syntagmatique et l'axe paradigmatique.

Le sens des mots dérivés dépend en partie de la nature de l'affixe. L'addition du préfixe *pré-* ajoute ainsi la signification d'une antériorité temporelle (TLFI) au mot auquel il est combiné : *préavis* (avant l'avis), *préhistoire* (avant l'histoire), *prédire* (annoncer à l'avance). Le suffixe *re-* ajoute la signification "faire à nouveau" comme dans *redire* (dire à nouveau quelque chose). Il faut cependant noter que le sens que certains affixes ajoutent à la racine n'est pas toujours explicite et que la connaissance des significations de chaque morphème n'est pas toujours suffisant pour comprendre le sens du mot polymorphémique (par ex. *-asse* a généralement une valeur d'amplification, néanmoins ceci n'est pas toujours clair. Par exemple dans le mot *vinasse*, *-asse* a plutôt une valeur péjorative). Sur ce point, la morphologie dérivationnelle diffère de la morphologie flexionnelle, cette dernière regroupant des suffixes beaucoup plus transparents sémantiquement que la première.



D'ailleurs, pendant longtemps, le cas des formes fléchies n'a pas suscité un grand intérêt dans le cadre des études portant sur la reconnaissance des formes polymorphémiques, car elles étaient considérées comme appartenant au domaine de la syntaxe et non au domaine lexical. Cependant, en considérant les cas d'irrégularités, les recherches actuelles accordent un regain d'intérêt à cette morphologie.<sup>5</sup>

### **Le statut des mots polymorphémiques**

Si les morphèmes peuvent être considérés comme des unités insécables et totalement arbitraires, cette notion d'arbitrarité est plus floue en ce qui concerne les mots polymorphémiques. En effet, ceux-ci étant composés de plus d'un morphème, leur sens est souvent déductible par l'analyse de ces derniers, et leur forme n'est donc plus tout à fait immotivée. Par exemple, le mot *lentement* est construit à partir des morphèmes *lent-* et *-ment*, le premier exprimant l'idée d'un manque de rapidité, le second servant à construire des noms d'action<sup>6</sup> ; en parlant d'un mouvement, *lentement* désigne donc une manière de faire les choses sans rapidité. Le sens des mots polymorphémiques n'est donc en général pas entièrement arbitraire si l'on se place du point de vue des morphèmes à partir desquels ils sont construits. Cette particularité des mots polymorphémiques peut conduire à se questionner sur la façon dont ils sont considérés par les locuteurs d'une langue. Sont-ils des mots dont le signe est indissociable de ceux des morphèmes qui les constituent, ou ont-ils un statut à part entière ?

Certains modèles linguistiques proposent, pour une langue donnée, un inventaire de toutes ses unités lexicales avec les principes généraux indiquant les règles lexicales (c'est-à-dire de formation de mots) qui opèrent sur l'ensemble de ces unités. Selon les modèles,

---

<sup>5</sup>Nous reviendrons de manière plus précise sur ces cas d'irrégularités dans le Chapitre 3.

<sup>6</sup>La plupart des suffixes ont une part d'interprétation très générale. Ce qui les différencie tient plus à leur origine, populaire ou savante et au type de la racine à laquelle ils sont adjoints. On peut distinguer quelques grandes interprétations : par exemple les suffixes *-age*, *-ment*, *-ion* servent à former des noms d'action ; les suffixes *-esse*, *-té*, *-isme* servent à former des noms abstraits ; les suffixes de nom ou d'adjectif *-et*, *-ette*, *-ot*, *-otte*, *-ard*, etc. ont une valeur appréciative (cf. Huot, 2001, p. 64-65).

ces unités lexicales correspondent aux morphèmes et aux mots polymorphémiques (théorie des unités entières), ou bien uniquement aux plus petites unités de sens - les morphèmes (théorie des unités réduites). Dans la théorie des unités entières, tous les mots (monomorphémiques ou polymorphémiques) sont des unités complètement spécifiques. Ainsi, Dell (1979) envisage le lexique comme un ensemble fini d'unités lexicales entièrement spécifiées (les lexies), accompagné d'un ensemble de règles (selon J-P Babin, 2000, p. 94). Dans la théorie des unités réduites, seuls les morphèmes sont considérés dans le lexique et les règles de formation des mots polymorphémiques sont localisées dans une composante à part : la composante syntaxique (cf. Chomsky, 1965). Au cours de cette thèse, nous reprenons cette thématique en faisant appel aux méthodes d'investigation issues de la psycholinguistique. Dans cette discipline, bon nombre d'"outils" théoriques sont issus de la linguistique ; néanmoins, la manipulation cognitive des mots est l'objet d'étude plutôt que la description des mots pour eux-mêmes. Nous allons donc à présent aborder la question de la faculté de langage chez l'homme du point de vue de la psycholinguistique, en nous centrant sur le problème que posent les mots polymorphémiques quant à la nature du signe qui les représente.

### **Les questions que pose la représentation des mots polymorphémiques au sein du lexique mental**

La faculté propre à l'homme de communiquer par l'intermédiaire d'un système de signes a conduit les psycholinguistes à postuler l'existence nécessaire d'un système lexical interne nous permettant, d'une part, de stocker en mémoire la forme de tous les mots de la langue que nous connaissons et, d'autre part, d'encoder le lien entre la forme et le sens de ces mots. La notion de lexique mental est la métaphore largement admise aujourd'hui par les psycholinguistes pour rendre compte de ce système.

*"En stockant la forme et l'information significative ensemble, le lexique (mental) résout le difficile problème de la correspondance arbitraire entre la forme et la signification"*  
(Frauenfelder, 1991, p. 7)

Le lexique mental serait ainsi une sorte de dictionnaire interne nous permettant d'accéder au sens de tous les mots d'une langue - à partir du moment où nous la manipulons suffisamment bien - de manière efficace. Toutefois, l'analogie avec un dictionnaire s'achève ici, car au vu de la rapidité avec laquelle nous manipulons une langue que nous avons bien intégrée, l'organisation des représentations des mots à l'intérieur de ce lexique n'emprunte probablement pas une structure linéaire comme l'ordre alphabétique d'un dictionnaire. Les mots polymorphémiques, dont nous discutons la notion d'arbitraire, posent aux psycholinguistes la question de la nature de leur représentation au sein du lexique mental. En effet, la question est de comprendre si ces mots ont leur propre représentation au sein du lexique mental, ou bien si seuls les morphèmes qui les composent sont représentés, et quelles sont les procédures mises en place pour accéder à ces représentations (pour illustrer ces questions voir la Figure 1.3).

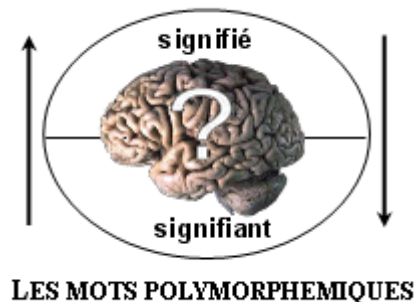


FIG. 1.3 – Quelle est la nature du signe des mots polymorphémiques pour les locuteurs d'une langue ?

En faisant l'hypothèse que les procédures de traitement du langage ne sauraient être indépendantes de la structure de la langue, il est acceptable de postuler que les représentations des mots sont organisées en fonction de leurs caractéristiques intrinsèques dans le lexique mental. Ainsi, certaines théories psycholinguistiques proposent qu'il existe au sein du lexique mental une représentation lexicale partagée du morphème dont sont composés les mots polymorphémiques d'une même famille morphologique (Taft et Forster, 1975).

En effet, il ne serait pas nécessaire de prévoir une représentation lexicale pour les mots *jardinier*, *jardinage* ou *jardinière* ect. puisque le sens de ces mots est directement dérivable du sens du morphème *jardin* et de celui de leur affixe *-ier*, *-age* ou *-ière*. Ainsi, les mots polymorphémiques ne possèderaient pas leur propre représentation, mais seraient encodés dans le lexique mental de manière “décomposée”. Seuls l’élément racine et les suffixes de ces mots auraient une représentation lexicale propre, mais pas leur combinaison. Par ailleurs, un travail réalisé par Rey-Debove (1984) permet d’estimer à 75 % les mots de la langue française analysables en constituants morphémiques. L’économie sur le nombre d’entrées lexicales opérée par un tel mode de stockage ne serait ainsi pas du tout négligeable. Néanmoins, et en contre-partie, une telle organisation implique que les mots polymorphémiques soient décomposés au cours de leur accès lexical, procédure qui est effectivement proposée dans l’hypothèse décompositionnelle concernant l’accès au sens de ces mots. Deux limites significatives ont été avancées concernant cette hypothèse. D’une part, le processus de composition-décomposition ayant un coût cognitif, l’économie faite en termes mnémoniques n’en est plus nécessairement une du point de vue du traitement global du langage. D’autre part, comme nous l’avons déjà mentionné un peu plus haut, certains mots polymorphémiques sont peu prédictibles quant à leur sens par l’analyse de leurs constituants, et posent donc un problème à l’hypothèse décompositionnelle. Henderson (1985) signale par exemple que le sens du préfixe anglais *un-*, bien que clairement négatif, peut apparaître sous différentes variantes sémantiques :

*“An unarmed man may never have been armed but an unfrocked priest has undoubtedly been de-frock-ed. An unearthed object has, at least metaphorically, been dug up, as if from the earth, whereas an unearthly object has never been part of the earth.”* (Henderson, 1985, p. 38).

En réponse à ces critiques, d’autres théories proposent que chaque mot ait sa propre représentation lexicale (Butterworth, 1983 ; Rumelhart & McClelland, 1986). Elles renvoient à l’hypothèse d’un accès direct aux représentations globales des mots polymorphémiques. Selon celles-ci, des connexions lexicales sont établies entre les mots qui partagent des

similarités orthographiques, sémantiques, voire morphologiques, et l'accès aux représentations lexicales d'un mot donné propage de l'activation aux représentations auxquelles il est connecté (Henderson, 1985). Une limite à ce type d'hypothèse est qu'un grand nombre de formes lexicales doit alors être stocké dans le lexique mental, et si celles-ci ne sont pas activées suffisamment souvent, elles "dépérissent". Pour palier ces limites, les modèles à accès direct comportent un mécanisme qui permet qu'à chaque fois qu'une représentation lexicale reçoit de l'activation, de manière directe ou par le biais de ses connexions, le seuil d'activation à partir duquel cette représentation peut être activée baisse. Ainsi, les seuils d'activation des entrées lexicales de tous les mots - même ceux qui ne sont pas fréquemment utilisés - sont maintenus à un niveau suffisamment bas pour permettre un accès à leur sens qui soit toujours rapide et efficace. Sans cela, les mots les moins fréquents deviendraient de plus en plus difficiles à activer et seraient graduellement oubliés.

L'objectif en psycholinguistique concernant l'étude des mots polymorphémiques est de déterminer dans quelle mesure une réalité linguistique descriptive, basée sur l'étude des énoncés linguistiques, constitue aussi une réalité en termes de processus de traitement du langage. L'idée de dissocier les unités morphémiques et les mots polymorphémiques est séduisante en terme organisationnel, cependant le sens des mots polymorphémiques est généralement un peu plus que la somme du sens des unités qui les composent. Dès lors, la question soulevée est de comprendre dans quelle mesure il est possible de dissocier la représentation lexicale ou formelle d'un mot de la représentation sémantique correspondante. Une idée en psycholinguistique pour résoudre ce dilemme a été d'envisager un traitement autonome de la forme des unités lexicales vis-à-vis de leur traitement sémantique. Ce type de raisonnement a conduit d'une part à des propositions de modèles de reconnaissance des mots polymorphémiques via une procédure de décomposition de ces derniers. A l'inverse, d'autres modèles ont proposé que les mots polymorphémiques soient traités par les locuteurs d'une langue via leur forme globale (hypothèse d'un accès direct). D'autres approches encore proposent un mécanisme de reconnaissance de ces mots basé à la fois sur la forme globale et sur les produits d'une décomposition. Nous allons expliciter

plus en détails ces catégories de modèles. Avant cela, et pour fournir tous les éléments nécessaires à leur pleine compréhension, nous détaillons quelques-uns des principaux modèles de reconnaissance visuelle des mots. Ces modèles peuvent être regroupés en deux familles qui reposent sur deux types distincts de mécanismes de reconnaissance visuelle des mots. A la suite de leur description, nous verrons que les modèles de reconnaissance des mots polymorphémiques reprennent préférentiellement à leur compte un type de mécanisme ou un autre, en fonction de l'hypothèse soutenue (décompositionnelle ou d'accès direct). Ce sont ces derniers points que les expériences que j'ai menées pendant ma thèse me permettront de discuter.

## **1.2 La reconnaissance visuelle des mots**

Quand nous lisons des mots qui constituent une phrase, nous prenons conscience uniquement du résultat de cette lecture. La facilité avec laquelle nous parvenons à traiter les mots écrits nous conduit généralement à considérer cette capacité comme un processus très élémentaire, celui d'un simple appariement d'une forme visuelle (grapho-orthographique) à un contenu sémantique. Le caractère rapide, irrépressible, et efficace de ce processus (c.f. effet Stroop, 1935) nécessite pourtant, selon toute vraisemblance, un système de traitement bien organisé dont différents modèles tentent de rendre compte à leur manière. Nous allons voir qu'un modèle phare de la reconnaissance des mots et celui du modèle à double voie de traitement (Coltheart, 1978). Des procédures plus fines pour le traitement des mots ont été également proposées en parallèle. Celles-ci se centrent plus particulièrement sur l'activation des représentations lexicales qui codent les mots au sein du lexique mental. Nous allons voir que selon les hypothèses défendues, l'activation de ces représentations peut solliciter des processus par recherche active ou par accumulation passive d'informations. Cette dernière hypothèse a donné lieu à la naissance d'un nouveau type de modèles de lecture dits connexionnistes, pour lesquels une seule voie est disponible et suffisante (McClelland & Rumelhart, 1981).

### 1.2.1 Modèles à double voie

Un exemple prototypique du modèle à double voie est celui de Coltheart (1978). Ce modèle postule l'existence de deux voies (ou procédures) parallèles et indépendantes permettant de traiter tout item pouvant être rencontré par un lecteur. De manière plus précise, les étapes de la lecture d'un mot au sein de ce modèle sont les suivantes : les informations visuelles qui correspondent à une séquence de lettres sont transmises depuis les récepteurs de la rétine au cortex visuel dans lequel un traitement de ces informations, impliquant d'autres régions cérébrales (pariétal), a lieu et permet d'identifier les lettres et de coder leur position. C'est à ce niveau que deux voies de lecture sont alors possibles.

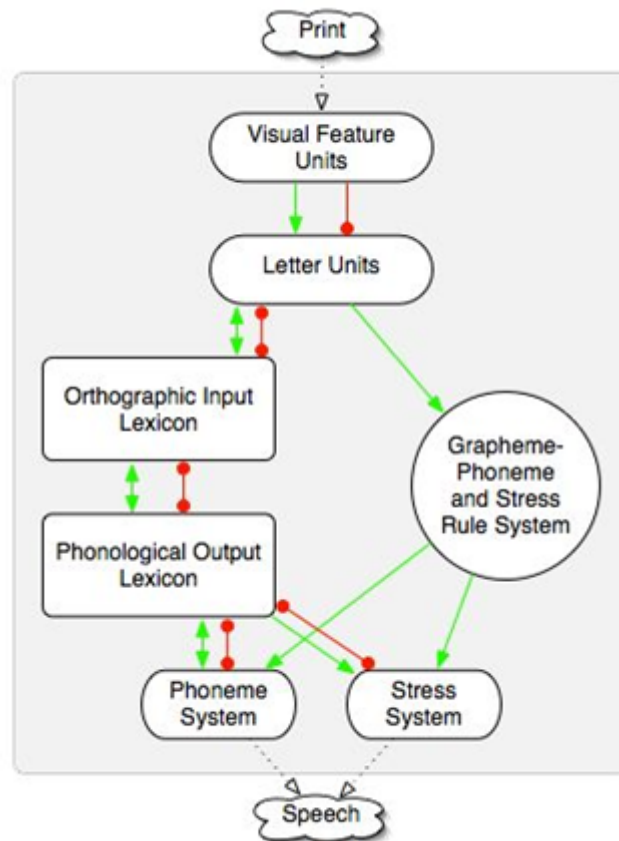


FIG. 1.4 – Version actuelle du modèle de la reconnaissance des mots à double voie : modèle *Dual Route Cascade* (DRC) d'après Coltheart et al. (2001).

1. La voie lexicale, qui va permettre d'activer automatiquement la représentation orthographique complète et spécifique du mot traité, puis sa représentation sémantique et phonologique.
2. La voie extra-lexicale, à partir de laquelle la série de lettres va être segmentée en unités graphémiques mises en correspondance avec des unités phonologiques, grâce à un ensemble de règles de correspondance graphème-phonème. Appliquées séquentiellement, ces dernières permettent d'accéder à la représentation sémantique du mot qui pourra éventuellement être prononcé.

Ces deux voies de traitement ont leur pertinence pour des types de mots distincts. La première voie, dite par adressage, est très efficace chez le lecteur compétent pour les mots familiers même irréguliers. La seconde voie, par assemblage, est indispensable pour traiter les mots nouveaux et les non-mots. Une version plus actuelle de ce modèle est proposée par Coltheart et al. (2001) avec le modèle *Dual Route Cascade*. Au sein de celui-ci, des interconnexions entre les niveaux des représentations phonologiques, orthographiques et sémantiques sont prévues. Par ailleurs, les deux voies de traitement (par adressage et assemblage) ne sont pas hermétiques (voir la Figure 1.4).

### 1.2.2 L'activation des représentations lexicales

Des procédures plus fines concernant l'activation de la représentation orthographique d'un mot ont été détaillées parallèlement à la proposition du modèle plus général de la double voie de traitement. Par exemple, Forster (1976) propose qu'une procédure de recherche active et séquentielle soit engagée dans le lexique afin d'activer la représentation orthographique d'un mot traité. Plus précisément, il propose que deux sous-systèmes entrent en jeu dans ce processus : un *système périphérique* et un *système central*. Ces deux systèmes se différencient par le fait que l'un, le système périphérique, est spécifique à une modalité alors que le second, le système central, est amodal. Au sein du système central sont stockées les représentations lexicales sémantiques et syntaxiques des mots ainsi que leurs représentations orthographiques ou phonologiques. Au sein du système



périphérique à modalité visuelle dans le cas de la lecture, les représentations sont de nature orthographique et servent essentiellement de code d'accès dans le mécanisme de reconnaissance des mots. Ainsi, lors de la reconnaissance d'un mot, la représentation orthographique de celui-ci est le code à partir duquel s'opère la recherche lexicale. Celle-ci se fait de manière sérielle sur un ensemble de représentations stockées dans le système central qui partagent des caractéristiques orthographiques avec le code. Quand la bonne représentation est contactée dans le système central, celle-ci est activée, et le stimulus mot reconnu : les informations sémantiques et syntaxiques qui lui sont associées deviennent disponibles pour le lecteur (voir la Figure 1.5).

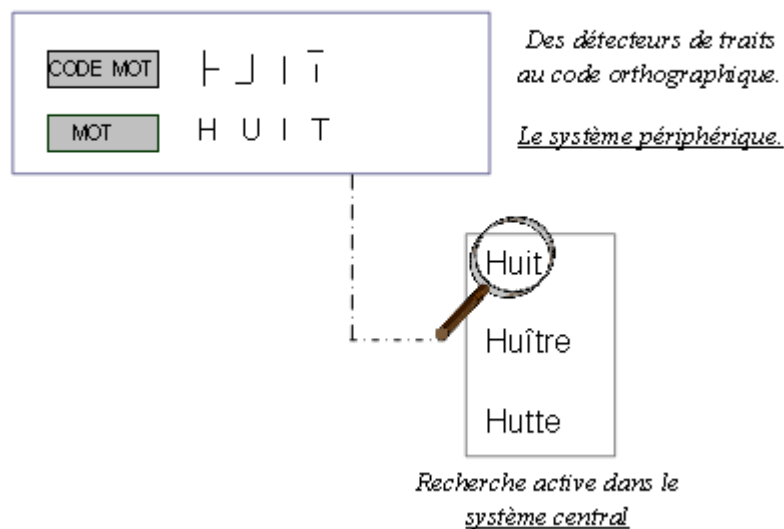


FIG. 1.5 – L'accès au lexique par recherche active. Mise en correspondance du code orthographique aux représentations encodées dans le fichier central.

Ce que nous pouvons noter ici est que dans ce type de modèle, la recherche de la représentation orthographique d'un mot pourrait être réalisée sur des unités plus réduites que la forme complète du mot (par ex. celles des syllabes ou des morphèmes). En contrepartie, ceci nécessite qu'une analyse pré-lexicale ait lieu avant cette étape de traitement.

## Rôle des unités morphologiques dans l'accès lexical

Taft et Forster ont proposé une étude princeps en 1976 qui montre que les unités morphémiques jouent un rôle dans l'accès lexical aux mots polymorphémiques. Cette étude, proposée en anglais, consiste en une tâche de décision lexicale sur des non-mots correspondants à des racines liées (par ex. *\*juvenate*, qui existe comme racine liée dans *rejuvenate*) et des non-mots "classiques" (par ex. *\*pertoire*, qui n'est pas une racine liée en anglais). Les auteurs tentent de montrer l'implication des unités morphologiques dans les processus d'accès lexicaux, et testent la conséquence la plus extrême de cette hypothèse, c'est-à-dire si les racines liées sont représentées dans le lexique mental. Leur prédiction est la suivante : si les racines liées sont représentées dans le lexique, alors les participants de l'étude devraient éprouver davantage de difficultés à les considérer comme des non-mots en comparaison des non-mots classiques. De manière plus précise, leur hypothèse est que les non-mots du type de *\*pertoire* devraient être immédiatement classés en tant que tels car aucune représentation correspondante ne peut être trouvée dans le lexique. En revanche, si l'on considère des non-mots qui correspondent à une racine liée (*\*juvenate*), alors une représentation lexicale devrait être contactée dans le lexique et renvoyer l'information selon laquelle cet item est un morphème lié, et donc pas un mot ; le système chercherait alors s'il n'existe pas une forme similaire qui corresponde à une racine libre<sup>7</sup>. Les résultats observés par les auteurs mettent en évidence des temps de réponse plus longs pour rejeter un non-mot comme *\*juvenate* comparé à un non-mot du type *\*pertoire*, ce qui suggère qu'une recherche lexicale a bien été engagée sur les non-mots qui correspondent à une racine liée et qui requièrent de ce fait plus de temps pour être rejetés en tant que mots.

Une deuxième expérience a été réalisée à la suite de celle que nous venons de présenter afin de lever toute ambiguïté concernant un biais possible induit par les participants. En effet, il se pourrait que ces derniers prennent plus de temps pour rejeter les non-mots qui correspondent à une racine liée parce qu'ils vérifient que ces items ne peuvent pas être

---

<sup>7</sup>Ceci est parfois le cas, comme par exemple *vent* ('bouche de conduit' en français) qui peut être un mot ou une racine liée dans *invent* ('inventer') ou *prevent* ('prévenir').

employés de manière isolée, sans pour autant qu'une procédure d'activation d'une représentation lexicale soit impliquée. Pour palier cet éventuel biais, la deuxième expérience proposée consiste en une tâche de décision lexicale sur des non-mots préfixés, par exemple *\*dejuvenate* et *\*depertoire*. Si le non-mot *\*dejuvenate* est plus long à rejeter que *\*depertoire*, cela montrera que de tels non-mots ont été décomposés, et que c'est la pseudo-racine isolée (*\*juvenate*) qui ralentit la décision de rejeter *\*dejuvenate* comme étant un mot, selon le principe de recherche lexicale expliqué précédemment. Dans le cas contraire, s'il n'y a pas de décomposition et donc isolation de la racine liée *\*juvenate*, alors *\*dejuvenate* et *\*depertoire* devaient être rejetés aussi rapidement l'un que l'autre car aucune représentation ne serait jamais activée dans le lexique mental. Encore une fois, les résultats observés vont dans le sens de l'hypothèse des auteurs, et sont donc en faveur d'une procédure d'accès lexical réalisée sur la base de représentations morphémiques. Nous reviendrons sur le rôle de ces unités lorsque nous présenterons les modèles de reconnaissance visuelle des mots polymorphémiques.

### **Rôle des unités syllabiques dans l'accès lexical**

Une autre unité qui aurait un rôle possible à jouer dans les processus de reconnaissance des mots est l'unité syllabique. Sphær et Smith (1973) montrent par exemple son rôle dans le traitement des mots écrits : ils observent que le pourcentage d'identification correcte de mots de même longueur (5 lettres) est plus élevé pour les mots contenant une seule syllabe (par ex. *paint*, 'peindre') que pour ceux en contenant deux (*paper*, 'papier'), ceux-ci suggèrent ainsi que l'unité syllabique pourrait intervenir dans la procédure qui conduit à la reconnaissance lexicale des mots. Les mots bisyllabiques seraient identifiés moins vite car non pas une, mais deux procédures de traitement susceptibles de conduire à la reconnaissance lexicale seraient mises en route : une pour chacune des deux syllabes. Un travail réalisé en anglais par Prinzmetal, Treiman et Rho (1986) a également montré que le traitement des mots écrits faisait appel à leur structure syllabique. Dans leurs travaux, ils observent en effet que le rappel de la couleur d'une lettre cible est facilité quand celle-ci

appartient à une syllabe qui est écrite d'une seule et même couleur dans un mot. Ainsi, la couleur de la lettre D sera mieux rappelée si elle a été présentée dans **VOD**/KA que dans **VO**/DKA (voir aussi Rapp, 1992 en anglais; Carreiras, Alvarez & De Vega, 1993 en espagnol; Colé, Magnan, Zagan & Grainger, 1999 en français). Une étude proposée par Fabre et Bedoin (2003) montre également l'importance de l'unité syllabique dans une étude similaire à celle de Prinzmetal et al. (1986), mais réalisée sur des non-mots. Ceci conforte l'idée que cette unité joue un rôle dans les traitements qui précèdent la reconnaissance d'un mot. De nombreuses autres études, menées notamment en langue française et utilisant d'autres procédures expérimentales que celles citées plus haut, confirment le rôle important de la syllabe dans le traitement des mots (citons par ex. les études de Melher & al., 1981; Ferrand et al., 1996; Pallier et al., 1993, menées en lecture ou en dénomination, ou de Colé et al., 1999 pour une tâche centrée sur l'écrit). Néanmoins, selon les langues, le rôle de l'unité syllabique n'est pas toujours aussi déterminant qu'en français. Par exemple, des unités infra-syllabiques semblent être aussi impliquées en anglais. Les études de Stemmer (1983) ou de Treiman et Danis (1988) montrent ainsi que l'attaque (c'est à dire la consonne ou le groupe de consonnes en position initiale dans la syllabe) et la rime (la voyelle et les éventuelles consonnes suivantes) sont des unités importantes dans le traitement des mots de cette langue. Par ailleurs, et comme nous allons le voir dans les sections qui suivent, d'autres unités semblent être sollicitées, sur lesquelles pourrait s'effectuer la recherche lexicale.

### **Rôle des unités phonologiques dans l'accès lexical**

Une expérience réalisée en 1987 par Van Orden en anglais montre que les représentations phonologiques ont une influence sur l'activation des représentations complètes des mots. En effet, ces auteurs ont montré que la reconnaissance du mot *rose* ('rose') est facilitée si celui-ci est précédé par la présentation du mot *rows* ('lignes'). Cette facilitation, observée alors que les deux mots ne partagent pas de sens mais uniquement leur forme phonologique, montre que des traitements d'unités de connaissance phonologique

sont mis en jeu et participent à la reconnaissance des mots en propageant de l'activation aux représentations de ces derniers. Nous ne décrivons pas plus les études qui portent sur cet effet (pour une revue, voir Berent & Perfetti, 1995 et Frost, 1998), et attirons plutôt l'attention sur le fait que le rôle de l'unité phonologique dans l'activation d'une représentation complète des mots a inspiré un nouveau type de modèle d'accès lexical, à savoir les modèles à voie unique ou connexionnistes.

### 1.2.3 Modèles à voie unique ou connexionnistes

Les modèles connexionnistes à traitement parallèle et représentations distribuées principes sont ceux de McClelland et Rumelhart (1981) et Rumelhart et McClelland (1982), et plus récemment celui de Plaut, McClelland, Seidenberg et Patterson (1996). Ces modèles reposent sur les mécanismes de fonctionnement des réseaux neuronaux, réseaux et mécanismes physiologiques qui sont repris et simplifiés dans les modèles informatiques dits connexionnistes. A l'inverse des modèles à double voie, ces modèles supposent qu'un système unique de lecture est disponible et suffisant. Plus précisément, leur architecture est celle d'un réseau à trois couches (la couche des unités d'entrée, celle des unités cachées, et celle des unités de sortie) interconnectées. Chaque unité est un "accumulateur" d'indices, et transmet un signal aux unités de la couche suivante. La valeur et la polarité des connexions sont déterminées par apprentissage, processus par lequel les valeurs des connexions se modifient graduellement sur la base des différentes répétitions d'activation. On parle ici de représentations distribuées pour rendre compte du fait que, dans ces modèles, il n'existe pas d'unité de représentation spécifique pour chaque mot ou lettre ; chaque unité intervient au contraire dans la représentation de plusieurs mots, et chaque mot est représenté par l'activation d'un réseau d'unités formant une configuration spécifique (voir Figure 1.6). Ce type de modèle a également été inspiré par le modèle dit passif de Morton (1969). Au sein de ce dernier, l'accès au lexique se réalise par le biais de détecteurs de mot (les *logogènes*) qui accumulent de l'activation en provenance du stimulus traité de manière passive.



FIG. 1.6 – Schéma du fonctionnement du modèle connexionniste de McClelland et Rumelhart (1981).

Les détecteurs qui reçoivent le plus d'activation sont ceux dont les unités lexicales auxquelles ils réfèrent partagent le plus de caractéristiques avec le stimulus en cours de traitement. Lorsque la valeur seuil prédéfinie d'un détecteur est atteinte, le mot correspondant peut être identifié. La différence majeure entre les modèles connexionnistes et le modèle de Morton est que les premiers postulent que les détecteurs existent pour des unités de tailles variées (traits, lettres, mots), étagées sur des niveaux en interaction. Par ailleurs, les modèles connexionnistes ne retiennent pas l'hypothèse d'une analyse pré-lexicale des mots en structures infra-lexicales. Cette dernière notion n'a en effet pas vraiment de sens puisque le traitement en cascade amène à stimuler des détecteurs de mots sur la base d'une activation partielle de lettres qui ne doivent pas nécessairement être toutes identifiées.

Dans les versions plus récentes de ces modèles, trois types d'unités seraient en fait simultanément impliquées dans le processus de reconnaissance de mots écrits (voir Figure 1.7) et contribueraient à reconnaître le mot (voir par ex. Ziegler, Van Orden et Jacobs, 1997). Dans la section suivante, nous allons pouvoir à présent décrire quelques modèles

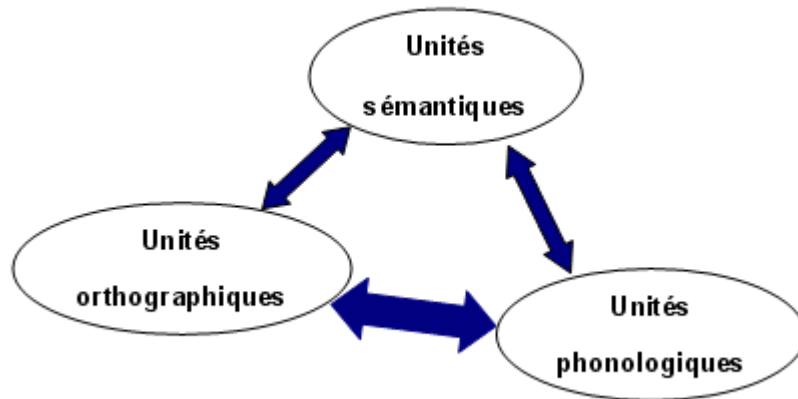


FIG. 1.7 – Schéma d'un modèle dans lequel trois types d'unités sont simultanément impliqués dans le processus de reconnaissance de mots écrits.

plus spécifiques de la reconnaissance visuelle des mots, à savoir ceux qui décrivent les processus d'accès lexical aux mots polymorphémiques et l'organisation de ces derniers au sein du lexique mental. Deux catégories de modèles peuvent être opposées en fonction des procédures de traitement lexical qu'elles postulent : soit le traitement se réalise via une procédure de décomposition des mots polymorphémiques en racine plus affixe(s), avec une recherche lexicale basée sur la racine, soit il se fait directement à partir de la forme globale du mot polymorphémique.

## 1.3 La reconnaissance visuelle des mots polymorphémiques

### 1.3.1 Modèle décompositionnel

Ce type de modèle a été formulé initialement par Taft et Forster en 1975 et Forster en 1976. Son principe de base est que les mots polymorphémiques sont décomposés en racine et affixe(s) au cours des traitements pré-lexicaux. Cette hypothèse décompositionnelle est

en faveur d'une organisation du lexique mental où seuls les éléments morphémiques possèdent une représentation spécifique dans les voies d'accès au lexique. Les mots polymorphémiques ne possèdent donc pas de représentations lexicales propres et indépendantes, et seules leurs racines sont stockées dans le lexique mental (ou fichier central) avec toutes les informations concernant les affixes avec lesquels elles peuvent se combiner pour former des mots polymorphémiques acceptables. Dans ce modèle, les mots polymorphémiques appartenant à une même famille morphologique partagent la même entrée lexicale dans la voie d'accès, qui est la racine morphémique commune (voir Figure 1.8).

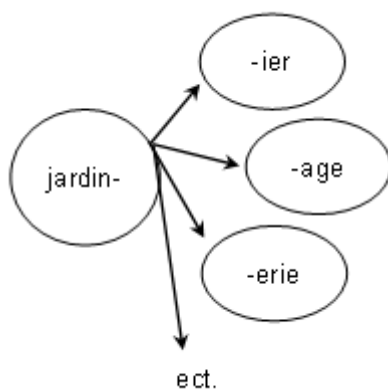


FIG. 1.8 – Schéma représentant le mode d'organisation des mots appartenant à une même famille morphologique dans le cadre du modèle décompositionnel.

### 1.3.2 Modèle à accès direct (ou à listage exhaustif)

Selon ce modèle, tous les mots connus par un sujet sont listés sous une forme unitaire et autonome dans le lexique mental (Butterworth, 1983). L'accès aux mots polymorphémiques ou monomorphémiques relève donc d'un même processus, à savoir d'un accès via la représentation globale du mot. La dimension morphologique n'est pas représentée en tant que telle, mais plutôt en termes de liens lexicaux (voir Figure 1.9).

Pour rendre compte de la capacité à créer de nouveaux mots polymorphémiques, ainsi que de l'intuition des locuteurs quant à l'appartenance à une même famille de mots com-



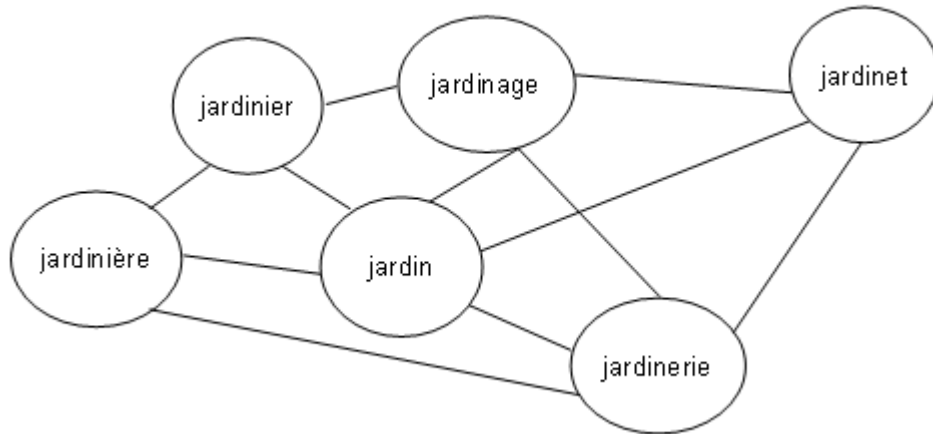


FIG. 1.9 – Schéma représentant le mode d’organisation des mots appartenant à la même famille morphologique dans le cadre du modèle à listage exhaustif.

posés à partir d’une même racine, ce modèle inclut des procédures dites supplétives (*fallback procedures*) consciemment utilisées par les individus. Des éléments de l’hypothèse du listage exhaustif se retrouvent dans les modèles connexionnistes (par ex. les mots polymorphémiques sont tous représentés sous leur forme holistique dans le lexique)<sup>8</sup>. Dans ces derniers, les capacités cognitives humaines sont considérées dans le cadre de réseaux d’unités lexicales qui peuvent être implémentés à l’aide de réseaux neuronaux artificiels. Les comportements cognitifs y sont le résultat d’interactions entre plusieurs de ces unités élémentaires du réseau. L’acquisition de connaissances ou d’habiletés cognitives se traduit par un processus d’ajustement du poids des connexions partagées entre certaines unités.

<sup>8</sup>Il est important de préciser ici qu’il est possible de scinder les modèles connexionnistes en deux grandes catégories : d’une part, les modèles dans lesquels la structure du réseau et le sens des unités neurones sont fixés à l’avance par les concepteurs, et d’autre part les modèles dans lesquels ces caractéristiques émergent à la suite d’un apprentissage (par ex. quand on règle les poids de perceptrons multicouches avec un jeu d’entrées-sorties d’apprentissage). Dans le cas des modèles psycholinguistiques de reconnaissances de mots, nous retrouvons majoritairement la première catégorie de modèles dans lesquels chaque forme holistique correspond à un neurone (comme dans la couche supérieure du modèle de McClelland & Rumelhart, Figure 5). Ceci permet de dire que ces modèles font appel à une forme de listage exhaustif.

Selon cette approche, les mots polymorphémiques d'une même famille morphologique entretiennent des connexions orthographiques/phonologiques et sémantiques privilégiées au sein du réseau. A chaque fois qu'un mot polymorphémique est activé, celui-ci propage de l'activation aux représentations qui partagent les informations qu'il contient, comme par exemple celles de sa racine et des autres mots composés à partir de celle-ci. Ces interactions consolident les liens entre ces représentations lexicales. Les informations morphologiques apparaissent dès lors comme des propriétés émergentes de ce système (Rumelhart & McClelland, 1986; Mac Whinney & Leinbach, 1991). En somme, les effets morphologiques qui peuvent être observés en lecture de mot (lecture plus rapide du mot *pur* si celui-ci est précédé par le mot *pureté*<sup>9</sup>) peuvent être attribués aux interconnexions sémantiques et orthographiques des mots d'une même famille morphologique.

### 1.3.3 Le débat entre modèle à décomposition et modèle à accès direct

#### Etudes sur les formes verbales régulières et irrégulières en anglais

Un débat existe entre les tenants des modèles dans lesquels les mots polymorphémiques sont traités via les représentations de leurs morphèmes et les mots non décomposables, via leur représentation globale (par ex. la *dual route hypothesis* de Pinker, 1991), et les tenants des modèles à procédure unique pour tous les mots, qui sont tous traités via leur forme globale (par ex. les modèles connexionnistes de Rumelhart & McClelland, 1986 et de Plunkett & Marchman, 1993). Pour départager ces deux conceptions du traitement lexical, de nombreuses études sur l'accès au sens des formes verbales régulières et irrégulières en anglais ont été réalisées. Dans cette langue, seuls les verbes réguliers sont décomposables (par ex. *talked*, qui est la forme passé du verbe *talk*, 'parler', est décomposable en *talk* et *-ed* qui marque le passé), tandis que les verbes irréguliers sont indécomposables;

---

<sup>9</sup>Nous revenons sur ce type d'expérience un peu plus bas.

leurs formes étant idiosyncratiques (par ex. *brought* est le passé de *to bring* ‘apporter’<sup>10</sup>). Selon les modèles à traitement indifférencié, l’accès lexical des formes fléchies régulières et irrégulières se fait via leur représentation globale. Selon les modèles à double voie, un mécanisme de type holistique est engagé uniquement pour le traitement des formes verbales irrégulières, tandis que les formes verbales régulières sont traitées sur la base d’un système d’application de règles : une racine plus une terminaison en *-ed* correspondent à la forme au passé d’un verbe régulier.

### Données en faveur des modèles à double voie

Des expériences de décision lexicale accompagnées d’un protocole d’amorçage en modalité visuelle masqué ou en inter modalité ont montré des effets d’amorçage sur la reconnaissance d’un verbe régulier (*walk*, ‘marcher’) quand celui-ci est précédé de la présentation de sa forme fléchie au passé (*walked*). En revanche, pour les verbes irréguliers cet effet d’amorçage n’est pas observé. La reconnaissance du verbe *give* (‘donner’) n’est par exemple pas facilitée quand celui-ci est précédé de la présentation de sa forme fléchie au passé *gave*<sup>11</sup> (Kempley & Morton, 1982 ; Napps, 1989 ; Stanner, Neiser, Herson & Hall, 1979 ; Marslen-Wilson, Hare & Older, 1995). Ces données sont interprétées comme la marque que les formes verbales régulières sont décomposées au cours de leur traitement lexical. De cette manière, l’amorce *walked* est décomposée en *walk* plus *-ed*, ce qui isole une première fois la représentation de la racine *walk* alors activée. Quand le participant doit traiter la cible *walk*, ceci est facilité, car sa représentation a été pré-activée au cours

---

<sup>10</sup>D’un point de vue purement linguistique, les verbes irréguliers anglais ne sont pas entièrement idiosyncrasiques. Halle and Marantz (1993) soulignent que dans le cas des verbes irréguliers, le choix du suffixes est soit *-Ø* (par ex. *hit* est la forme passée du verbe *hit* ‘frapper’) soit *-t* (par ex. *slept* est la forme passée du verbe *sleep*, ‘dormir’ adjoint à la forme phonologique alternative du radical, qui est déterminé par l’identité du verbe et spécifié dans une liste stockée en mémoire.

<sup>11</sup>Lorsque le temps de présentation de l’amorce est très rapide ou lorsqu’elle est présentée en modalité auditive, le lien sémantique partagé entre l’amorce et la cible dans les exemples donnés ne permet pas d’obtenir d’effet facilitateur sur la reconnaissance de la cible ; le traitement de l’amorce a été interrompu avant que sa représentation sémantique ait été entièrement activée.

du traitement de l'amorce. Les formes verbales irrégulières en anglais n'étant pas décomposables et leur forme idiosyncrasique n'étant pas composée de leur racine, le traitement de l'amorce (*gave*) ne permet pas de faciliter celui de la cible (*give*) : il n'y a pas d'effet d'amorçage observé.

### **Données issues de la neuropsychologie cognitive en faveur du modèle à double voie**

Des études sur des patients atteints de lésions cérébrales localisées dans le lobe frontal ou souffrant de la maladie de Parkinson ont montré des difficultés dans le traitement des formes verbales régulières (Tyler, Mornay, Anokhina, Longworth, Randall & Marslen-Wilson, 2002; Ullman, Corkin, Copolla, Hickok, Growdon, Koroshetz & Pinker, 1997). A l'inverse, des patients atteints de lésion cérébrales dans le lobe temporal inférieur présentant la démence de type d'Alzheimer montrent des difficultés dans le traitement des formes verbales irrégulières (Kay, Lesser & Coltheart, 1992; Ullman & al., 1997). Cette correspondance entre des zones d'atteinte cérébrale différentes et des déficits cognitifs distincts pour le traitement des formes verbales régulières et irrégulières (double dissociation) est en faveur des modèles à double voies de traitement.

Plus précisément, Ullman & al., (1997) argumentent que le cortex peut être sub-divisé, du point de vue des processus mnésiques, en deux systèmes : la mémoire déclarative, qui sous-tend la mémorisation de faits et d'évènements, et la mémoire procédurale, qui est le système pour mémoriser des savoir-faire et des habiletés cognitives. La mémoire déclarative est prise en charge par le lobe temporal et pariétal du cortex cérébral, et la mémoire procédurale par le lobe frontal (Cohen & Squire, 1980). Étant donné que les mots sont considérés comme des unités correspondant à des faits et non des savoir-faire, ils seraient mémorisés dans la mémoire déclarative. En revanche, les règles grammaticales seraient elles mémorisées dans la mémoire procédurale. Pour tester l'hypothèse de la double voie (voie de traitement lexical par application de règles et la voie par accès direct), les auteurs proposent de tester des patients atteints de la maladie d'Alzheimer ou de Parkinson

dans une tâche de production de formes verbales régulières ou irrégulières. Sachant que les patients atteints d'Alzheimer ont des difficultés au niveau de la mémoire déclarative, mais des compétences préservées pour la mémoire procédurale, si le modèle à double voies est valable, alors ces patients devraient être plus aptes à produire des formes verbales régulières qu'irrégulières en comparaison de sujets sains. Les patients Parkinsoniens présentent le schéma inverse, à savoir que leur mémoire déclarative est préservée, mais qu'ils manifestent des troubles concernant leur mémoire procédurale. De ce fait, si le modèle à double voie est valable, les patients Parkinsoniens devraient mieux produire les formes verbales au passé des verbes irréguliers que celles des verbes réguliers. Les résultats observés vont dans le sens des prédictions faites. Cette double dissociation est en faveur du modèle à deux voies de traitement, à savoir une voie qui applique des règles grammaticales pour les verbes réguliers, et une voie qui permet d'activer directement en mémoire la représentation des formes verbales irrégulières.

### **Interprétation par les modèles connexionnistes**

Les modèles connexionnistes interprètent les effets d'amorçage morphologique comme des propriétés émergentes du système lexical et des interactions qui s'y opèrent entre les unités lexicales. En effet, dans ces modèles, les mots polymorphémiques ont leur propre représentation et celles-ci sont interconnectées entre membres d'une même famille morphologique par des liens établis sur des proximités phonologiques et sémantiques. Ainsi, lors de l'activation de la représentation d'un mot polymorphémique, cette activation est propagée aux autres membres de la famille qui partagent tous la même racine. Celle-ci est alors activée et permet d'obtenir les effets d'amorçage observés dans les tâches de décision lexicale avec protocole d'amorçage. Concernant les résultats de Ullman et al. (1997), la dissociation observée entre les patients Alzheimer et Parkinsoniens pourrait s'expliquer dans un modèle connexionniste en termes de dommage des représentations de nature sémantique ou phonologique. En effet, étant donné que la forme fléchie au passé d'un verbe irrégulier est idiosyncrasique (par ex. *gave*), il faut, pour être en mesure de retrouver celle-

ci dans une tâche de production, que le lien de nature sémantique avec le verbe auquel elle se rattache soit intact - pour *gave* il s'agit d'un lien avec *give*. Concernant les verbes réguliers maintenant, dont les formes passées sont phonologiquement très proches de la racine, celles-ci seront moins bien traitées si les représentations de nature phonologique sont endommagées <sup>12</sup>.

En conclusion, ces deux modèles, l'un à décomposition pré-lexicale pour les mots polymorphémiques et l'autre à accès direct pour tous les mots, constituent deux points de vue opposés quant à l'organisation du lexique mental. Un autre débat concernant le traitement lexical des mots polymorphémiques porte sur le rôle des représentations morphémiques et globales au cours de l'accès lexical. Des modèles à deux voies de traitement sur les mots polymorphémiques (aussi appelé modèles mixtes ou de compromis) ont en particulier été proposés. Ils postulent que les mots polymorphémiques sont traités via une procédure de décomposition et par un accès direct. La question se pose dès lors pour ces modèles de déterminer laquelle des procédures conduit la première à la reconnaissance.

## 1.4 Les modèles de compromis

Le modèle de Taft et Forster (1975) est un modèle princeps en psycholinguistique. Il décrit les mécanismes qui sous-tendent l'accès lexical aux mots polymorphémiques et qui sont aussi à la source des modèles de compromis. Leur modèle s'appuie sur l'hypothèse forte selon laquelle seuls les composants morphémiques des mots polymorphémiques possèdent une représentation lexicale (voir Figure 1.10). Par exemple, pour le mot *revenir*, la représentation lexicale à partir de laquelle le système va retrouver les informations lexicales correspondantes est celle de la racine *venir*. Etant donné le format dans lequel sont stockés les mots polymorphémiques, leur traitement lexical doit nécessairement se faire via une procédure de décomposition (par ex. décomposition en *re - venir* pour le

---

<sup>12</sup>Voir l'article de Joanisse & Seidenberg de 1999, dans lequel une simulation connexionniste reproduisant les processus décrits ci-dessus produit les mêmes schémas de résultats que ceux observés par Ullman & al., (1997) avec des patients Alzheimer et Parkinsoniens.

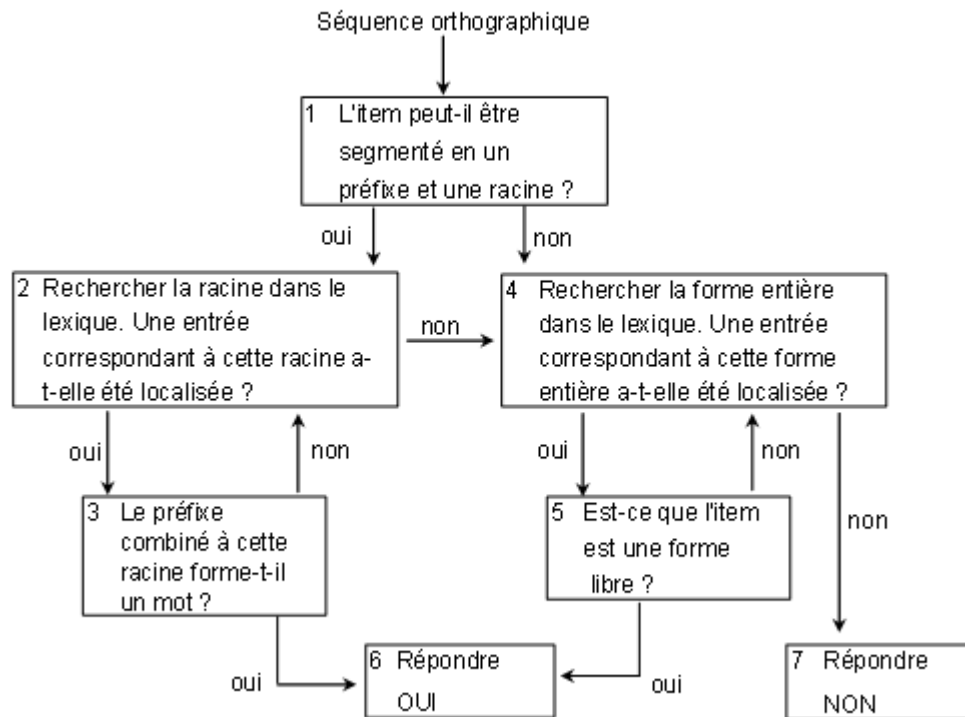


FIG. 1.10 – Modèle de Taft & Forster (1975) rendant compte des procédures engagées au cours d’une tâche de décision lexicale sur des mots polymorphémiques (d’après Taft & Forster, 1975, p. 644).

mot *revenir*). Une fois les représentations des composants morphémiques activées, une étape de vérification est mise en route pour vérifier la légalité de la combinaison de l’affixe et de la racine. Si cette procédure aboutit, les informations lexicales du mot sous sa forme intégrée (*revenir*) sont rendues disponibles. Si une des étapes n’aboutit pas (étape de décomposition, de vérification...), alors le système s’oriente vers une autre procédure qui vise à retrouver une forme globale du mot traité dans le lexique. Par exemple, pour un mot comme *décorer*, la procédure de décomposition par le processus d’*affix stripping* (‘dépouillage de l’affixe’ en français) conduit à l’isolation des unités *dé-* et *-corer*, alors que cette dernière forme n’est pas une racine en français et n’a donc pas de représentation dans le lexique mental. De ce fait, une recherche additionnelle est lancée, basée sur la

forme entière du mot. Dans le cas du mot *démanger*, la procédure de décomposition isole les morphèmes *dé-*, *mang-* et *-er*. La racine *mang-* est bien représentée dans le lexique, mais la procédure de vérification n'aboutit pas, car la représentation de cette racine n'encode pas l'information selon laquelle elle peut se combiner au préfixe *dé-*. Une procédure de recherche additionnelle est ici à nouveau déclenchée sur la base de la forme globale du mot *démanger*. Une limite que l'on peut tout de suite retenir à l'encontre de ce modèle est l'augmentation de la complexité du traitement - dans les cas où les procédures de décomposition et de vérification n'aboutissent pas. Un modèle qui propose un traitement par décomposition et/ou par accès direct dès les premières étapes du traitement lexical a été proposé pour palier ces limites (voir le modèle à compromis AAM présenté ci-dessous).

#### 1.4.1 L'Augmented Address Model (AAM) de Laudanna et Romani (1985)

L'*Augmented Address Model* (AAM) de Laudanna et Romani (1985) est un modèle qui comporte deux niveaux entre l'entrée visuelle du mot à traiter et sa représentation sémantique (voir pour des modèles similaires Caramazza, Laudanna & Romani, 1988; Burani & Laudanna, 1992; Caramazza, Miceli, Silveri & Laudanna, 1985; Laudanna, Badecker & Caramazza, 1992; et Frauenfelder & Schreuder, 1992 ou Schreuder & Baayen, 1995). Ces deux niveaux correspondent à celui des unités d'accès et celui des représentations lexicales. Les unités d'accès peuvent coder la forme entière des mots polymorphémiques (accès unitaire) ainsi que leur forme décomposée (accès morphologique). Les représentations lexicales orthographiques (dans le cas de la lecture) sont quant à elles sous un format décomposé, avec les racines représentées séparément des affixes (voir Figure 1.11). Au niveau du fonctionnement, ce modèle propose que les deux procédures d'accès (unitaire et morphologique) fonctionnent en parallèle quand un mot polymorphémique est présenté. La procédure d'accès qui atteint le plus rapidement un niveau d'activation prédéfini active les représentations sous le format décomposé au niveau supérieur (lexique orthographique d'entrée). Dans ce modèle, les mots polymorphémiques fréquents sont reconnus via la



représentation de leur forme globale. Dans le cas des mots polymorphémiques peu fréquents ou rencontrés pour la première fois, l'unité d'accès qui permet d'accéder au lexique orthographique est celle codant les unités morphémiques.

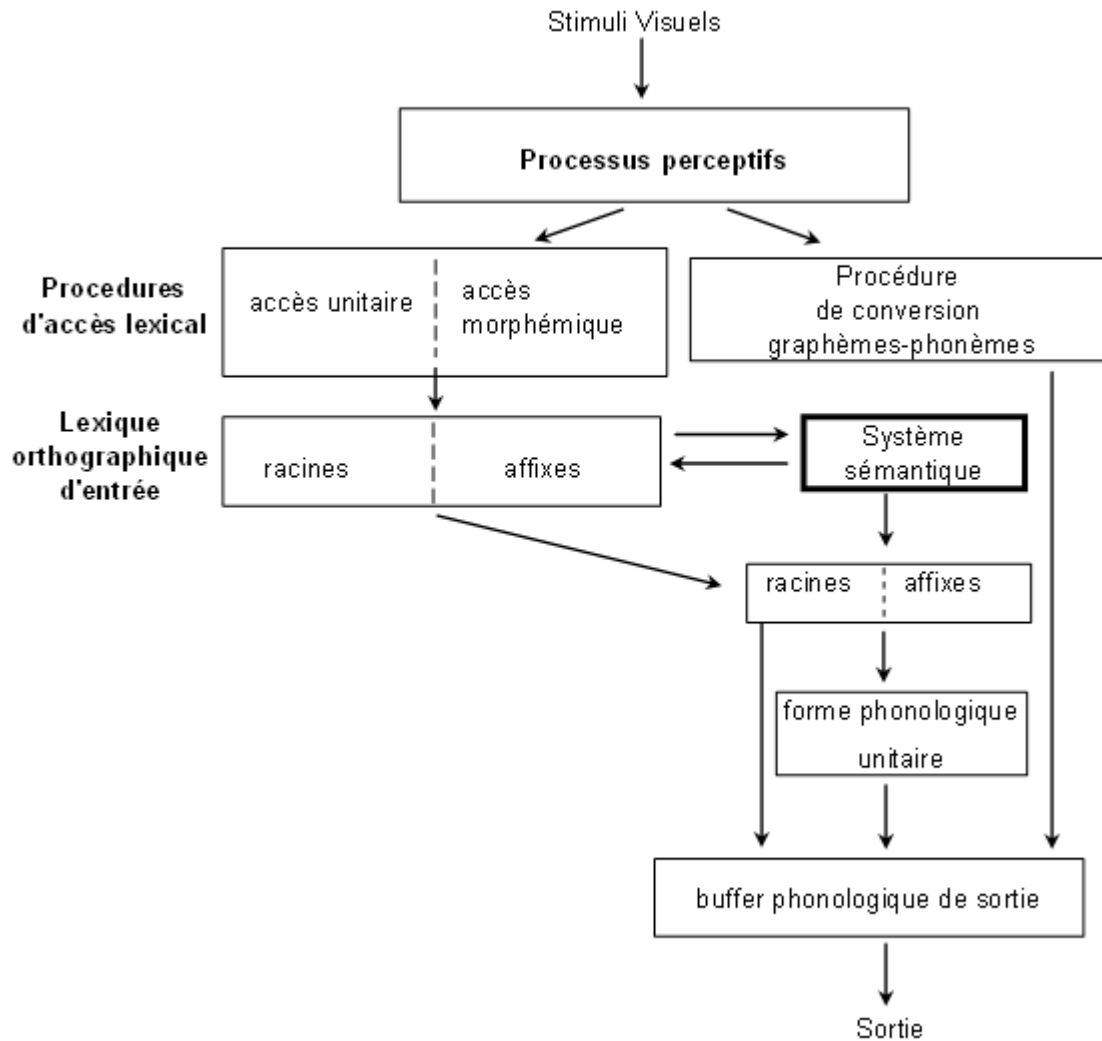


FIG. 1.11 – Schéma du modèle AAM, d'après Laudanna et Romani (1985).

### 1.4.2 Le modèle de Taft (1994)

Afin de répondre à la critique concernant la complexité qu'entraînait son premier modèle de reconnaissance des mots polymorphémiques, Taft a proposé un second modèle.

Celui-ci, présenté dans un article de 1994, est composé de trois niveaux de représentations qui interagissent entre eux : un niveau morphologique, un niveau lexical et un niveau conceptuel. Au niveau morphologique sont représentés les affixes et les racines liées ou libres. Au niveau lexical apparaissent les formes globales et les morphèmes libres. Les informations sémantiques sont disponibles au niveau conceptuel. Dans ce modèle, le traitement d'un mot polymorphémique sémantiquement transparent (*revenir*) se fait par la décomposition précoce de celui-ci, permettant d'activer les représentations de *re-* et *-venir* au niveau morphologique.<sup>13</sup> Les représentations au niveau morphémique permettent l'activation de leur représentation lexicale (au niveau supérieur) et la signification du mot global correspond à l'activation conjointe de ces représentations conceptuelles. Dans le cas des mots polymorphémiques sémantiquement opaques, comme par exemple le mot *démanger* (qui ne signifie pas 'arrêter de manger'), la procédure de décomposition active la représentation de la racine *manger* et du préfixe *dé-* aux niveaux morphémique et lexical. La représentation au niveau morphémique de *manger* est interconnectée à la représentation lexicale de *manger* et *démanger*. L'affixe *dé-*, quant à lui, n'est pas connecté à la représentation de *manger* mais uniquement à celle de *démanger*, qui reçoit dès lors plus d'activation que *manger* et active à son tour et plus fortement sa représentation globale au niveau conceptuel.

### Données expérimentales en faveur des modèles AAM et de Taft (1994)

Les deux modèles présentés ci-dessus émettent des hypothèses différentes quant au traitement de pseudo-mots. Le modèle de Taft, qui est un modèle à décomposition précoce, prédit que dans une tâche de décision lexicale, des pseudo-mots qui ont une structure morphologique illégale (nous appelons cette condition MI), comme en italien le pseudo-mot *cantevi* (avec *cant-* et *-evi* qui sont bien deux morphèmes en italien mais ne peuvent

---

<sup>13</sup>En fait, les affixes sont supposés être représentés dans un niveau syntaxique séparé du niveau morphologique. Néanmoins, pour expliquer les procédures de traitement, nous pouvons admettre que ces derniers sont représentés avec les représentations des racines.

apparaître ensemble puisque *cant-* est un verbe du premier groupe et *-evi* une terminaison du deuxième groupe) sont plus longs à rejeter que des pseudo-mots qui n'ont pas de structure morphologique comme *canzovi* en italien (nous appelons cette deuxième condition NM pour désigner une structure non morphologique). En effet, selon le modèle de Taft, un mot comme *cantevi* est décomposé par une procédure d'*affix stripping* et la racine *cant-* est alors isolée puis utilisée pour effectuer la recherche d'une entrée lexicale. Une fois l'entrée activée, la procédure de vérification teste s'il existe une entrée lexicale correspondant à la forme globale intégrant la racine et le suffixe. A ce niveau-ci du traitement, les pseudo-mots avec une structure MI sont rejetés dans une tâche de décision lexicale. Les pseudo-mots NM sont eux rejetés beaucoup plus rapidement car aucune entrée lexicale n'a pu être activée dès les premiers traitements, ce qui évite ainsi l'étape suivante de vérification. Le modèle de Taft prédit également que les pseudo-mots construits avec une racine existante et une terminaison qui n'est pas un suffixe (condition que nous nommons AR, pour avec racine), comme le pseudo-mot *cantovi* en italien, ne devraient pas engager de procédure d'*affix stripping*. Ainsi, ils devraient être rejetés au même niveau de traitement que les pseudo-mots NM. Enfin, des pseudo-mots construits avec une racine qui n'en est pas une en italien et un suffixe existant dans cette langue (condition AS, pour avec suffixe comme le pseudo-mot *canzevi*), devraient être rejetés plus lentement qu'un pseudo-mot AR. En effet, ce dernier engage une procédure lexicale par le biais de sa racine qui possède une représentation au sein du lexique mental, alors que cette procédure n'est pas engagée par un pseudo-mot AS qui contient une racine non représentée dans le lexique. Pour résumer, les temps de réponse prédits par le modèle de Taft concernant les pseudo-mots présentés ci-dessus s'ordonne de la façon suivante pour une tâche de décision lexicale :  $MI > AS > AR = NM$ .

Concernant maintenant le modèle AAM, au sein duquel une représentation globale et une représentation morphémique sont activables pour accéder au sens des mots polymorphémiques, les pseudo-mots activent des représentations lexicales orthographiquement similaires à leur forme (puisque'ils n'ont pas de représentation lexicale propre). Si un mor-

phème existant est contenu dans le pseudo-mot traité, alors sa représentation est activée. Les pseudo-mots MI et NM vont donc activer dans le lexique orthographique certaines unités lexicales qui leur sont similaires. Le pseudo-mot MI (*cantevi*) activera en particulier la représentation de la racine *cant-*. De ce fait, le modèle prédit que le temps mis pour rejeter les pseudo-mots MI sera plus long que pour les pseudo-mots NM, car d'avantage de représentations auront reçu des activations (même si le seuil d'activation est insuffisant pour déclencher la reconnaissance d'un mot). Concernant maintenant les pseudo-mots AR et AS, ils devraient activer la représentation d'unités qui leur sont orthographiquement similaires, ainsi que le morphème dont ils sont composés (la racine ou le suffixe). Cependant, la suite de lettres restantes ne trouve pas de représentation correspondante au niveau supérieur et de ce fait ces stimuli devraient être rejetés avant qu'une activation lexicale ait pu avoir lieu, donc avant les pseudo-mots de type MI. En résumé, le modèle AAM prédit l'ordonnancement suivant concernant les temps de réponse à ces pseudo-mots dans une tâche de décision lexicale :  $MI > AR = AS > NM$ . Ces prédictions sont testées par Caramazza et al. (1988). Les résultats obtenus (voir le Tableau 1.1) sont favorables aux modèles AAM et à celui de Taft (1994). En effet, le principal résultat est que les non-mots de type *cantevi* (morphologiquement illégal) ont des temps de réponses significativement plus long que pour tous les autres stimuli. Ceci est donc bien en faveur de modèles qui postulent que la structure morphologique des mots a un rôle important dans une tâche de décision lexicale, hypothèse contestée par les modèles à accès direct. Concernant les résultats obtenus pour les non-mots avec une structure morphologique partielle (AR et AS), les temps de réponses sont significativement plus longs pour les stimuli de type *canzevi* (AS) en comparaison des temps de réponse observés pour les stimuli qui n'ont pas de structure morphologique (NM). Maintenant, concernant les temps de réponses sur les stimuli du type *cantovi* (AS), les temps de réponses ne sont pas significativement différents de ceux observés pour les stimuli qui n'ont pas de structure morphologique (NM). Cependant, les résultats sur les taux d'erreurs montrent d'une manière significative davantage d'erreurs pour la condition AS que dans la condition NM. Ainsi les résultats de cette expérience

sont dans l'ensemble favorables au modèle AAM et à celui de Taft (1994), et les résultats ne permettent pas de véritablement trancher entre les deux.

	MI <i>cantevi</i>	AS <i>canzevi</i>	AR <i>cantovi</i>	NM <i>canzovi</i>
RT	<b>875</b>	<b>809</b>	<b>781</b>	<b>760</b>

TAB. 1.1 – Tableau résumant les principaux résultats observés dans l'étude Caramazza, Laudanna, & Romani (1988).

#### **Autres données expérimentales en faveur des modèles AAM et de Taft (1994)**

Un reproche qui peut être formulé à l'encontre des travaux qui portent sur des non-mots est que ces derniers, n'ayant pas de représentation lexicale, poussent le système à déclencher une procédure décompositionnelle, qui accroît la possibilité d'observer les effets attendus dans les modèles à décomposition. Un protocole expérimental permettant de tester la procédure décompositionnelle sur de vrais mots d'une langue est celui de la décision lexicale sur des mots polymorphémiques avec manipulation de leurs fréquences de surface et cumulée. La fréquence de surface correspond à la fréquence d'occurrence du mot polymorphémique considéré comme un item lexical isolé (par ex. la fréquence d'occurrence de *chanteur*). Cette mesure reflète la possibilité qu'à chaque fois qu'un mot polymorphémique donné est traité, sa représentation globale est activée. La fréquence cumulée correspond quant à elle à la somme des fréquences de tous les mots qui propagent une même racine (par ex. la fréquence de surface du mot *chant* + la fréquence de surface du mot *chanteur* + la fréquence de surface du mot *chanson* etc. . .). Cette mesure reflète l'hypothèse selon laquelle à chaque fois qu'un mot polymorphémique composé d'une racine donnée est traité, la représentation de sa racine est activée. A l'aide de ce protocole, Burani, Salmaso et Caramazza (1984) ont testé la reconnaissance de verbes réguliers en italien. Leur résultats

montrent que les deux fréquences précédentes influencent toutes deux les temps de reconnaissance. Ceci est interprété dans le cadre du modèle *Augmented Address Morphology* (AAM) grâce aux deux procédures de traitement des mots polymorphémiques dans les premiers niveaux du traitement lexical : la procédure décompositionnelle renvoie à l'effet de la fréquence cumulée et la procédure globale à l'effet de la fréquence de surface. Une autre interprétation pour expliquer la co-occurrence des effets des fréquences de surface et cumulée sur les temps de reconnaissance des mots polymorphémiques est possible au sein du modèle de Taft (1994). Selon ce modèle, l'effet de la fréquence cumulée rendrait compte de la procédure de décomposition (activation des représentations morphémiques) et l'effet de la fréquence de surface de la phase de recombinaison de la racine avec l'affixe en fin de traitement.

### 1.4.3 Le modèle de Giraudo et Grainger (2000)

Dans le modèle supralexicale de Giraudo et Grainger (2000), il existe deux niveaux de représentation entre l'entrée visuelle et les représentations sémantiques. Le premier niveau est celui des représentations orthographiques, (format global) le second celui des représentations morphémiques (format décomposé). La présentation d'un mot polymorphémique en entrée du système active une série de représentations correspondant à des descriptions orthographiques et phonologiques du mot. Ensuite, parmi ces représentations, celles qui correspondent à un ou plusieurs morphèmes sont connectées au niveau des unités morphémiques qui contiennent les affixes et les racines (voir Figure 1.12). Ce modèle fait l'hypothèse qu'à chaque fois qu'un mot polymorphémique ou qu'une racine libre est présentée à l'entrée du système, leurs représentations orthographiques envoient de l'activation aux représentations morphémiques auxquelles elles sont reliées. Celles-ci activent à leur tour l'ensemble des unités lexicales (orthographiques, phonologiques et sémantiques) auxquelles elles sont reliées. Par conséquent, dans ce type de modèle, les unités morphémiques que l'on pourrait qualifier de centrales permettent d'organiser les représentations orthographiques en familles morphologiques grâce aux connexions entretenues au sein de

ces dernières.

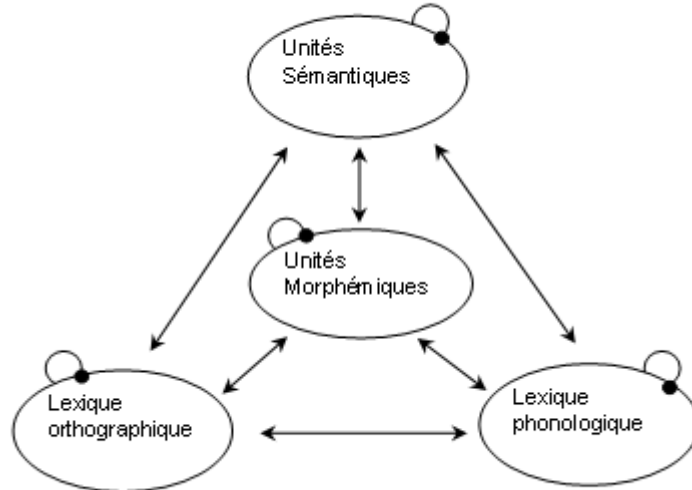


FIG. 1.12 – Schéma représentant le modèle supra-lexical proposé par Giraudo et Grainger (d’après Giraudo & Grainger, 2000).

### Données expérimentales en faveur du modèle de Giraudo et Grainger (2000)

Giraudo et Grainger (2000) observent que des mots dérivés de fréquence élevée (par ex. *amitié*) amorcent la reconnaissance de leur racine (*ami*) en comparaison d’une condition contrôle orthographique (*amidon*). Cet effet d’amorçage n’est pas observé quand la reconnaissance de la racine se fait après la présentation d’un mot dérivé de basse fréquence (*amiable*). Ces résultats ne vont pas en faveur des modèles sub-lexicaux qui prédisent que tous les mots polymorphémiques sont traités via une procédure de décomposition (*amiable* en *ami -able*), aboutissant à l’isolation de leur racine (*ami*) dont la représentation reçoit des activations. Cette préactivation permet dès lors à la racine d’être reconnue plus rapidement uniquement si elle est présentée seule dans une tâche de reconnaissance, et ceci indépendamment de la fréquence de surface du mot polymorphémique en amorce. Au contraire, dans un modèle supra-lexical comme celui de Giraudo et Grainger (2000), les

mots polymorphémiques sont traités sous leur forme globale, qui propage ensuite de l'activation à leurs représentations morphémiques. Dans le cas d'un mot polymorphémique de basse fréquence, l'activation propagée à sa racine est plus faible que celle propagée par un mot polymorphémique de fréquence élevée. Ce type de modèle permet donc de rendre compte des effets observés, à savoir un effet d'amorçage sur la reconnaissance d'une racine quand celle-ci est précédée d'un mot dérivé de fréquence élevée, et propageant dès lors suffisamment d'activation à sa racine pour qu'elle soit préactivée et reconnue plus rapidement.

#### 1.4.4 Le Parallel Dual Route Model (PDR) de Bertram, Baayen et Schreuder (2000)

Le modèle Parallel Dual Route de Bertram, Baayen et Schreuder (2000) est très proche du modèle AMM. Plus particulièrement ce modèle propose un approfondissement de la nature des procédures lexicales mises en jeu lors de l'accès au sens des mots polymorphémiques. Les auteurs décrivent trois étapes au cours du traitement lexical des mots. Au cours de la première, l'entrée visuelle active des représentations qui sont stockées dans la mémoire à long terme. Celle-ci contient une représentation du mot dans son format global, ainsi que les représentations des morphèmes qui le composent. Par exemple, un stimulus d'entrée comme le mot *dogs* ('chiens' en français) active les représentations des formes *dogs*, *dog*, *do* et *-s*. Les représentations parmi celles-ci qui dépassent un certain seuil d'activation prédéfini sont dirigées vers la mémoire à court terme et la deuxième étape du traitement débute. Au cours celle-ci, seuls les morphèmes plus courts que le stimulus mot traité restent activés. Cette procédure permet que la combinaison de morphèmes sélectionnée soit aussi longue que le stimulus mot traité et grammaticalement correcte. Enfin, au cours de la troisième étape, les caractéristiques syntaxique et sémantique des représentations lexicales qui sont encore activées se combinent permettant d'activer le sens du mot global. Selon la nature des mots polymorphémiques, la représentation globale (qui fait partie des représentations activées à la présentation du stimulus mot) est celle qui



permet l'accès au sens; dans d'autres cas, c'est la représentation morphémique. Etant donné que dans ce modèle, les mots polymorphémiques activent leur représentation globale et morphémique, le modèle de Schreuder et Baayen est considéré comme un modèle de compromis, avec deux voies de traitement qui n'entrent pas en compétition de manière directe mais qui travaillent en parallèle.

### Données expérimentales en faveur du modèle PDR

Schreuder et Baayen (1997) et New, Brysbaert, Segui, Ferrand et Rastle (2004) ont testé les temps de reconnaissance de noms écrits sous leur forme singulier ou pluriel. Ces mots sont soit :

1. *Singulier dominant*, c'est-à-dire que leur forme au singulier est plus souvent rencontrée que celle au pluriel (par ex. le mot *épouse*);
2. *Pluriel dominant*, c'est-à-dire que leur forme est plus souvent rencontrée au pluriel qu'au singulier (par ex. *nuages*).

Les temps de réponses collectés pour les mots présentés au pluriel sont plus courts lorsque le mot traité est *pluriel dominant* par rapport à ceux qui sont *singulier dominant*. Pour les mots *singulier dominant* les temps de reconnaissance sont plus courts quand les mots sont présentés au singulier et plus longs quand ils sont présentés au pluriel. Pour les mots *pluriel dominant* il n'y a pas de différence dans les temps de reconnaissance des mots qu'ils soient présentés au singulier ou au pluriel.

Ces résultats montrent que les mots écrits au pluriel et qui sont *pluriel dominant* (*nuages*) sont traités via leur forme globale, puisque les temps de réponses sont plus courts que pour un mot *singulier dominant* (comme le mot *épouses*). Les mots écrits au pluriel qui sont *singuliers dominant* sont reconnus quant à eux via la voie par décomposition et les temps de reconnaissance dépendent alors de la fréquence cumulée de la racine (fréquence qui correspond à la somme des fréquences de tous les mots polymorphémiques composés de la même racine *plus* un coût dû à la procédure de décomposition). C'est pourquoi un mot *singulier dominant* présenté sous sa forme singulier et qui est traité via

sa représentation globale est reconnu plus vite que lorsqu'il est présenté sous sa forme au pluriel. Dans ce dernier cas, il y a le coût de la procédure de décomposition qui rentre en compte et allonge le temps de reconnaissance. Le fait qu'il n'y ait pas de différence entre le temps mis pour reconnaître un mot *pluriel dominant* présenté au singulier et présenté au pluriel s'explique par le fait qu'il n'y a pas de temps pris par une procédure de décomposition pour le traitement de la forme *pluriel dominant* au pluriel car elle est traitée sous sa formes globale.

Les derniers modèles que nous venons de présenter proposent tous que les deux types de représentations - global et décomposé - entrent en jeu au cours de l'accès lexical aux mots morphologiquement composés. Une question centrale que posent ces modèles est la place relative de l'une et l'autre des deux représentations (morphémique ou décompositionnelle) dans l'accès lexical. Nous étudions cette question dans le Chapitre 2, en étudiant l'organisation et le traitement des mots dérivés en fonction de leur fréquence de surface. Dans le Chapitre 3, nous testons si une procédure de décomposition est mise en place même lorsque le traitement lexical concerne des formes verbales irrégulières en français. Dans les Chapitres 4 et 5, nous nous intéressons enfin de manière privilégiée au facteur qui déclenche la procédure de décomposition et si une structure neuronale particulière la prend en charge. L'ensemble de ces études met à profit la richesse linguistique de la langue française dans une grande variété de protocoles expérimentaux, permettant d'enrichir ainsi les approches qui ont été proposées jusqu'à présent.