

ROLE DE LA PRODUCTIVITE MORPHOLOGIQUE DANS LA RECONNAISSANCE DE LEXEMES COMPLEXES

Fanny Meunier, UMR 5596, CNRS & U. Lyon 2
fanny.meunier@univ-lyon2.fr

Georgette Dal, UMR 8163, CNRS, U. Lille 3 & U. Lille 1
georgette.dal@univ-lille3.fr

Bernard Fradin, UMR 7110, CNRS & U. Paris 7
bernard.fradin@linguist.jussieu.fr

Nabil Hathout, UMR 5610, CNRS & U. Toulouse le Mirail
nabil.hathout@univ-tlse2.fr

Abstract

In this paper we investigated the productivity of the suffix –ET in French. First, we did a linguistic description of derived words ending in *–et* and *–ette* that allowed us to identify different derivational patterns. This information gave us the active zones of the process and a first picture of its inscription in the language. We then made up neologisms on the basis of the patterns identified. These neologisms were then used to run two behavioural experiments looking at different perception of the different derivational patterns. Results revealed different patterns of results regarding derivational patterns used to create neologisms. We interpreted these results in terms of cognitive availability of the derivational patterns.

Résumé

Nous nous sommes intéressés à la productivité de la suffixation en –ET en français. Dans un premier temps, nous avons réalisé des descriptions linguistiques des dérivés en *–et* et *–ette* nous permettant d'identifier différents patrons de dérivation. Nous avons ainsi obtenu des informations sur les zones actives du procédé ainsi qu'une première image de son inscription dans la langue. Nous avons ensuite construit des néologismes sur la base des patrons identifiés. Ces néologismes nous ont permis de réaliser deux expériences visant à répondre à la question : y a-t-il une perception différenciée suivant le type de patron utilisé ? Les résultats mettent en évidence des patterns différents selon les patrons de construction des néologismes. Nous interpréterons ces résultats en termes de disponibilité cognitive des différents patrons morphologiques.

La psycholinguistique du mot s'intéresse aux représentations et aux mécanismes qui nous permettent de produire et d'interpréter les mots d'une langue. Afin d'expliquer comment l'homme traite le langage, il est nécessaire d'envisager l'existence d'un sous-système lexical – qu'il soit autonome ou pas – qui stockerait tous les mots et/ou formes de lexèmes¹ de la langue connus par un individu ainsi que les informations phonologiques, orthographiques et sémantiques liées : le lexique mental, ou lexique interne, métaphore largement admise aujourd'hui en psycholinguistique et de plus en plus convoquée par les morphologues théoriciens.

Le concept de lexique mental recouvre l'idée d'un système organisé des connaissances susceptibles de permettre à l'utilisateur un accès très rapide et efficace aux informations lexicales qui concernent des dizaines de milliers de mots. Si cette métaphore du lexique mental est assez consensuelle, différents problèmes, dont un d'importance fondamentale, restent sujets à débat : comment ce lexique interne est-il organisé en mémoire, c'est-à-dire quels sont la nature et le mode d'organisation des représentations mentales ?

L'unité de stockage dans le lexique mental pourrait être une unité plus petite que le mot. En effet, l'idée d'un listage exhaustif de toutes les unités lexicales connues par le sujet paraît peu économique du point de vue mnémotique. En revanche, une approche dans laquelle les régularités lexicales sont représentées par des règles et où seuls les éléments ne pouvant être établis par ces règles seraient listés, paraît plus fonctionnelle. Cependant il est nécessaire de tenir compte de l'accroissement de la complexité computationnelle qu'entraîne un système qui aurait à calculer, pour les rendre accessibles, les codes d'accès associés aux représentations lexicales.

Certaines théories soutiennent que les mots morphologiquement complexes doivent être décomposés en leurs constituants plus petits (les morphèmes des théories linguistiques structuralistes) avant l'accès au lexique (Taft & Forster, 1975), d'autres proposent que chaque forme existant dans une langue a une entrée lexicale propre et qu'il n'y a aucune analyse morphologique préalable à l'accès au lexique (Manelis & Tharp, 1977). Suivant l'option choisie, les structures morphologiques des mots en reconnaissance sont soit totalement transparentes, soit totalement opaques. Des positions moins tranchées sont également avancées dans la littérature récente (voir par exemple le méta-modèle proposé par Frauenfelder & Schreuder, 1992).

A notre connaissance, aucun modèle psycholinguistique ne fait intervenir la notion de productivité des patrons morphologiques permettant la formation des mots complexes dans le domaine de la reconnaissance de ces mots. Ce n'est que dans le domaine de la production que cette distinction apparaît. Dans ce domaine, la nature du mécanisme permettant la formation des mots complexes est considérée par certains auteurs comme une dimension pouvant affecter la structure du lexique mental, ainsi que les procédures qui s'y déroulent (Stemberger, 1985 mais voir Dell, 1986). De même, aucun modèle de reconnaissance des mots complexes n'évoque la disponibilité des procédés morphologiques. Le concept de disponibilité trace une distinction entre les divers procédés morphologiques de construction des mots construits. Sont dits disponibles (ou productifs au sens de Schultink, 1962) ceux que les locuteurs utilisent pour construire des dérivés non attestés. Les autres sont définis comme non disponibles.

¹ Les lexèmes correspondent au mot hors emploi, ils en sont une forme abstraite dépourvue de marques flexionnelles.

Nous pensons que ce type de distinction peut avoir une importance cruciale sur la nature des représentations lexicales associées aux mots complexes et leur étude peut nous permettre d'inférer les principes d'organisation interne des familles morphologiques. Nous avons concentré notre attention sur la suffixation en *-ET*.

Dans un premier temps, nous avons réalisé une description linguistique des dérivés en *-et* et *-ette*. La suffixation en *-ET* constitue le principal procédé de formation des diminutifs en français contemporain et le seul pleinement productif. Nous n'envisagerons ici que la dérivation des noms, qu'ils soient issus de noms (*cloche* > *clochette*), d'adjectifs (*bas* > *basset*) ou de verbes (*ronfler* > *ronflette*), à l'exclusion de celle des adjectifs (*gentil* > *gentillet*), qui pose des problèmes spécifiques que nous n'aborderons pas. Le choix d'étudier les mots suffixés en *-ET* du point de vue de la productivité se justifie par le statut ambivalent que présente ce suffixe sous ce rapport. D'une part, la suffixation en *-ET* est un procédé apparemment très productif en français, ce qui se manifeste par la création de nombreux types nouveaux (*macaronette*, *torsette*, *aspirette*) ; d'autre part, un nombre considérable de suffixés de ce type sont opaques (à des titres divers) et ne se rattachent à aucun patron actuellement productif.

Dans un travail antérieur nous avons entrepris de mesurer la productivité de la suffixation en *-ET* à partir du corpus du quotidien *Libération* couvrant les années 1995 à 1999. Cette étude constitue la première étude de la productivité de cette suffixation menée sur un corpus journalistique au moyen des outils de mesure basés sur le nombre d'hapax mis au point par H. Baayen (Max Plank, Nijmegen). Cette méthode utilisée dans de nombreux travaux sur la productivité morphologiques (Baayen & Renouf, 1996; Baayen & Lieber, 1991; Plag, Dalton-Puffer, & Baayen, 1999), définit la mesure de la productivité (dans un corpus) P comme :

$$P = \frac{n_1}{N}$$

où n_1 est le nombre d'hapax construits par un procédé morphologique donné et N le nombre total d'occurrences des formes de lexèmes construits par ce même procédé.

Dans un premier temps, nous présenterons les règles permettant de construire des dérivés en *-ET* (pour une présentation détaillée voir Fradin, Hathout et Meunier, 2003), que nous appelons aussi patrons. Nous présenterons ensuite deux expériences comportementales visant à évaluer à quel degré ces patrons sont disponibles pour les locuteurs. Nous mettrons enfin en rapport les résultats observés avec les mesures de productivité préalablement établies. Rappelons qu'un patron est une mise en correspondance réglée de variations entre le plan du substrat (phonologie, et partant graphie) et les plans sémantique et catégoriel. La caractérisation des différents patrons de construction est une étape nécessaire car aucun patron de dérivation unique ne peut fournir simultanément toutes les interprétations attestées.

1. Analyse descriptive

1.1. Deux pôles

Notre analyse défend l'idée que les suffixés en *-ET* du français sont organisés autour de deux grands pôles (Référent, Locuteur) et met en jeu une quinzaine de patrons de dérivation (pour les détails, voir Fradin *et al.*, 2003, Fradin, 2003). Le pôle Référent regroupe des significations ayant trait aux propriétés constitutives ou fonctionnelles de l'entité dénotée par le dérivé en *-ET*. Toutes les significations primaires de ce pôle mettent en jeu la notion d'échelle. L'effet sémantique induit est de l'ordre de la diminution : en gros, l'entité dénotée

par le dérivé possède une propriété à un degré inférieur au degré standard, qui vaut pour le référent du lexème-base. Le pôle Locuteur de son côté regroupe les dérivés dont la signification donne une information sur le rapport entre le référent du dérivé et le locuteur. Ainsi les dérivés du premier pôle véhiculent un contenu mesuratif (grand/petit), et ceux du second un contenu appréciatif (bon/mauvais).

1.2. Description des patrons

1.2.1. Patrons du pôle Référent

Dans tous ces cas, la suffixation en -ET indique que le degré d'une dimension que possède le référent du lexème dérivé est inférieur au degré que présente l'objet de référence, en l'occurrence le référent type du N-base. Nous avons dans ce cadre identifié dix patrons de dérivation, notés de R1 à R10.

En R1, un lexème du type *clochette* dénote un objet qui a une taille inférieure à celle d'une cloche, laquelle est censée avoir la taille standard pour les cloches. Autres exemples de R1 : *balisette, brosette, brochurette, châtelet, maisonnette...*

En R2, la situation est plus compliquée car plusieurs propriétés du référent sont prises en compte simultanément. On retrouve ici les cas pour lesquels le référent du dérivé présente sous une forme édulcorée les (ou certaines des) propriétés caractéristiques du référent du lexème-base : soit ces propriétés ne se retrouvent pas dans le référent du dérivé, soit elles s'y retrouvent sur un mode atténué e.g. *opéra / opérette, cigare / cigarette, sucre / sucrette*. Le cas R1 est un sous-cas de R2. Si on ne l'a pas fait figurer sous ce dernier, c'est à cause du nombre des types qu'il recouvre et de son importance comme prototype dans le dispositif cognitif de l'évaluation.

En R3, la dimension concerne exclusivement la sexuisemblance. R3 regroupe les significations où l'on bascule d'une valeur à l'autre : mâle vs femelle e.g. *merlette, fliquette, punkette, fantômette, beurette*. Ces dérivés présentent fréquemment une connotation hypocoristique (L6).

En R4, la propriété centrale est l'âge. L'interprétation est du type 'jeune de l'espèce Nb' e.g. *brebiette, porcelet*. Ce cas pourrait être vu comme une extension de R1 dans la mesure où il existe une implication générale disant que le jeune d'une espèce est de plus petite taille que l'adulte de la même espèce.

La dimension centrale pour R5 est le prix. Plusieurs de ces dérivés dénotent des imitations de fourrure dont la valeur est bien moindre que la fourrure authentique : *chinchillette, castorette, angorette*. Mais d'autres dénotent des produits de mercerie qui imitent le référent du nom base : *satinette, cotonnette, bisette* (Delhay 1996). On aurait aussi pu choisir de faire de cette signification un sous-type de R2 puisque les référents de R5 sont des ersatz du référent du N-base. La seule différence est que, dans ce dernier cas, le côté ersatz est sciemment revendiqué parce qu'il rend ces produits plus accessibles au chaland. Pour cette raison, on retrouve également les dérivés de type R5 dans le pôle Locuteur (L1).

R6 regroupe des dérivés dont le N-base dénote des individus ou des entités qui, par définition, exercent une fonction ou sont dotés de pouvoirs. La suffixation -ET indique que ces pouvoirs ou ces fonctions sont amoindris et n'ont plus les effets escomptés. *Roitelet* et *bombinette* en sont des exemples.

La fonction discriminée par R7 correspond à celle d'une substance servant d'ingrédient non majoritaire dans la préparation culinaire dénotée par le dérivé, tout en y étant saillante e.g. *anisette*, *vinaigrette* (Dal 1997 : 142).

R8 rassemble les dérivés qui dénotent des lieux où l'on effectue l'activité exprimée par le verbe-base et qui sont spécialement destinés à cet usage. Le suffixe -ET indique que l'activité en question est temporaire (vs. permanent). Le caractère temporaire est net dans les paires *cachette* / *cache* et *couchette* / *couche*.

Le type R9 dénote des événements ou des actions dont une ou plusieurs dimensions (durée, amplitude, intensité) sont diminuées par rapport à celles que manifeste le verbe base. Ces dérivés sont assez nombreux : *giclette*, *reniflette*, *risette*, *grimpelette*, *causette*, *trempelette*, *bronzette*.

Le type R10 enfin rassemble les dérivés qui dénotent des événements ou des actions dont l'effet est moindre que celui assuré par le référent du nom base : *réformette*, *mesurette*.

1.2.2. Patrons du pôle Locuteur

Nous avons dans le pôle Locuteur identifié 6 patrons de dérivation, notés de L1 à L6.

Le groupe L1 en général regroupe les dérivés dénotant des objets ou des parties de dispositifs conçus pour servir à l'accomplissement d'une fonction précise (artefacts fonctionnels) : *allumette*, *baïonnette*, *boulet*, *collet*, *serpette*, *mesurette*, *herminette*₃, *hachette*, *raclette*, *roulette*, *fourchette*, *douchette*, *chaufferette*, *lunette*, *mesurette*, *gâchette*, *pipette*, *mitraille*, *mouillette*, *moulinette*, *sonnette*, *traceret*, *arielette*, *vizirette*, etc. Rien d'étonnant à ce que beaucoup soient des outils ou des ustensiles. L'idée est qu'ils sont des objets maniables, dont il est facile de se servir ou qui facilitent l'exécution de certaines tâches. On peut discriminer des sous-groupes à l'intérieur de L1, pour lesquels l'apport de sens de la suffixation peut être précisé en fonction du type de l'objet. Celui des véhicules au sens large (*camionnette*, *fourgonnette*, *patinette*, *pétrolette*, *avionnette*, *caninette*) serait 'facile à conduire / manoeuvrer' ; celui des parties de vêtement (*épaulette*, *talonnette*, *cuisse*, *manchette*, *poignet*) serait quelque chose comme 'améliore (la fonction d') un vêtement'. Celui enfin des fourrures (*chinchillette*, etc.), déjà vues au pôle Référent en tant qu'imitations, pourrait être '(plus) faciles à acquérir' pour signifier qu'elles peuvent plus facilement être achetées et portées, etc. que les vraies fourrures.

En L2, le référent dénoté par le dérivé est quelque chose qui n'aurait pas existé sans l'action du référent du N-base, lequel est toujours une personnalité politique. L'effet interprétatif associé à L2 est celui d'une mesure administrative qui apporte un bénéfice inattendu dans le quotidien, même s'il est modeste. Ceci apparente clairement L2 à L6. Le procédé L2 est récent et productif, comme en témoigne la série *chalandonnette*, *baladurette*, *juppette*, *périssolette*, *arthuisette*, *aubryette*, *jospinette*.

Les dérivés L4 et L3 nous placent au cœur de la valeur d'appropriation reconnue à la suffixation -ET (Dal, 1997). Celle-ci exprime l'accointance qui existe entre le locuteur et le référent du dérivé. Elle se traduit par la manière même dont le dérivé est construit : il prend comme base un lexème exprimant une caractéristique physique ou comportementale saillante de ce référent. Mais cette caractéristique n'est saillante qu'à celui qui a une connaissance précise des particularités physiques de la plante ou de l'animal que dénote le dérivé ou de son mode de vie. Par exemple, dénommer *bergeronnette* un oiseau parce qu'il fréquente les berges des rivières (ou les bergeries) pour y trouver des moucheron suppose qu'on est familier de cet oiseau. Il s'agit d'une dénomination qui n'est pas purement descriptive. Les

dérivés de L3, L4 se situent aux antipodes des noms savants, qui sont élaborés sur la base de caractères observables et estimés pertinents dans la mise en place d'une taxinomie. Les caractéristiques les plus fréquentes en L3 et L4 sont la couleur (*rouget, bleuet*), le goût (*doucette*), l'odeur (*muguet* (< *musc*)), la forme (*rosette, pommette*), la nourriture (*chardonneret*: oiseau friand de chardons), l'habitat (*bergeronnette*), la saison d'apparition (*pâquerette* : fleurit vers Pâques) et la taille (*basset*).

Les dérivés centraux de L3 partagent les mêmes propriétés que ceux de L4. La seule différence est qu'ils dénotent des entités nuisibles ou mauvaises. Dénommer celles-ci au moyen du diminutif, c'est faire comme si elles appartenaient à la sphère intime du locuteur alors même qu'il a toutes les raisons de s'en garder. D'où la lecture propitiatoire associée à ces dérivés. L'exemple le plus clair est celui de *belette* (< *belle*).

Les dérivés L5 ont tous affaire au goût et à l'alimentation. Certains sont anciens comme *andouillette, sanguette, tartiflette*. Beaucoup sont récents et dénotent des aliments d'une variété nouvelle (*garriguettes, torsadette, macaronette*) ou d'un apprêt nouveau (*croquette, goujonnette, gigolette, jambonnette*). Le fait que leur nom soit une dénomination diminutive vise à suggérer que ces produits relèvent de la sphère des aliments les plus proches du goût des locuteurs, de ceux qui ont leurs préférences.

Les dérivés L6 sont des hypocoristiques. C'est là une fonction ancienne de la suffixation en -ET qui est toujours productive e.g. *fripounette, cochonnette, pépette, pulpinette*.

Nous avons dans Fradin et al. (2003) fourni une mesure de la productivité pour chaque type de dérivés en -ET distingué dans le classement ci-dessus (c'est-à-dire pour chacun des patrons de règle). A un niveau global, cette première étude a montré que la productivité de -ette est en général supérieure à celle de -et, ce qui justifie de ne pas considérer a priori que ces deux suffixes sont les exposants d'un même procédé morphologique. A un niveau plus fin, ce point de vue s'est trouvé renforcé par le fait que les patrons de dérivation pour lesquels -et et -ette sont productifs ne sont pas les mêmes. De plus tous les patrons dérivationnels ne sont pas productifs au même degré : certains paraissent improductifs (R8), d'autres sont très productifs (L6). Ce constat plaide pour ne pas traiter la suffixation en -ET comme un procédé unique. Ces résultats doivent être considérés comme indicatifs, car, du fait de la petite taille du corpus, ils reposent souvent sur des échantillons trop faibles pour être représentatifs.

A la suite de ce travail se posait la question de savoir si les locuteurs distinguaient le type du patron utilisé. Afin d'apporter des éléments de réponse à cette question, nous avons mis en place un protocole d'expérimentation. Nous avons choisi de restreindre cette première tentative au suffixe -ette, celui-ci étant plus productif que -et. Nous avons commencé par construire des néologismes sur la base des patrons de dérivation identifiés. Nous avons donc créé des mots nouveaux en combinant une base et le suffixe -ette. Il est à noter que certains de ces néologismes sont attestés dans les dictionnaires mais leur emploi est si rare qu'on peut postuler qu'ils n'ont jamais été rencontrés par la grande majorité des individus testés, et c'est ce critère qui est retenu dans une approche cognitive. Dans un premier temps, nous avons réalisé une expérience utilisant une tâche de décision lexicale, nous permettant de récupérer des temps de réaction reflétant le temps nécessaire pour rejeter le néologisme comme mot mais également un taux d'erreur dans le cas où le néologisme serait accepté comme mot. Dans une deuxième expérience, nous avons demandé à des étudiants d'explicitier le sens des néologismes que nous avons créés.

2. Expérience 1

Dans cette première expérience, nous avons demandé aux participants de décider le plus rapidement possible et en faisant le moins d'erreurs possible si l'item qui leur était présenté était un mot de la langue française ou non. Cette tâche de décision lexicale, classique dans les travaux de psycholinguistique, peut sembler délicate tant l'interprétation d'« être ou ne pas être un mot de la langue française » peut varier selon les individus. Effectivement, on peut interpréter cette consigne comme signifiant 'est-ce que l'item appartient à une liste préétablie ?' (la liste de référence restant à définir, i.e. la liste des mots tolérés à l'école, celle des dictionnaires académiques –voir l'étude de Corbin, 1987, à ce sujet -, ou encore la liste des mots utilisés quotidiennement), ou encore comme signifiant 'est-ce que l'item est susceptible d'être utilisé quand on parle le français?'. Clairement, les différentes interprétations mènent à la réalisation de tâches cognitives différentes telles que l'exploration mémorielle ou bien la génération de contextes d'utilisation. Pour réduire cette ambiguïté et minimiser les effets métalinguistiques, nous avons ajouté à cette première consigne des contraintes de temps et de qualité (« veuillez répondre le plus rapidement possible et en faisant le moins d'erreurs possible »). Les réponses observées se situent ainsi autour de 800 millisecondes (ms) avec un pourcentage d'erreur moyen de 9% (incluant les erreurs motrices), suggérant que, globalement, la tâche ne pose pas de problème particulier aux participants. Il semble que, dans ce cadre, l'interprétation effectuée par les participants de la consigne est de type 'est-ce que j'ai déjà vu ce mot?'. Il faut noter ici que le système cognitif est particulièrement performant à déterminer s'il a déjà vu/entendu quelque chose, et c'est en grande partie ce sentiment de familiarité qui va déterminer la réponse 'mot' ou 'non-mot' du sujet. De toute façon, quelle que soit la réponse donnée, elle est prise en compte soit dans l'analyse des temps de réponse obtenus à la réponse juste, soit dans l'analyse des taux d'erreurs. Il faut également garder à l'esprit que c'est moins le type de réponse qui est intéressant ici que les différences observées pour les différents patrons testés.

2.1 Méthode

2.1.1 Matériel.

Huit patrons ont été choisis pour servir de modèles à la fabrication de néologismes. Ils sont repérés ici par un des dérivés typiques qu'ils permettent de construire :

- 1- R1 et R2 : *clochette, opérette* ;
- 2- R3 : *merlette* ;
- 3- R5 et L1f : *castorette* ;
- 4- R7 : *vinaigrette* ;
- 5- R8 : *couchette* ;
- 6- R9 : *grimpette* ;
- 7- L2 : *balladurette* ;
- 8- L5 : *jambonnette*.

Entre 6 et 11 néologismes de chaque catégorie ont été sélectionnés et testés (voir annexe). Les participants devaient effectuer une tâche de décision lexicale : des items leur étaient présentés sur un ordinateur, l'un après l'autre. Pour chacun d'entre eux, le sujet devait appuyer le plus vite possible et en faisant le moins d'erreurs possible sur un bouton si l'item était un mot de la langue française et sur un autre bouton si l'item n'était pas un mot. Afin d'avoir tous les types

de réponses, en plus de nos néologismes expérimentaux, nous avons ajouté à notre liste 120 lexèmes en *-ette* existants de haute et de basse fréquences comme *maisonnette* et *voilette*, 52 néologismes de remplissage ainsi que 60 pseudomots non décomposables de type *calopette*.

2.1.2 Composition des listes expérimentales.

Les 300 items étaient vus par chacun des participants. L'ordre d'apparition des items était aléatoire.

2.1.3 Procédure expérimentale.

Les sujets ont passé l'expérience individuellement dans une pièce prévue à cet effet. Les items étaient présentés visuellement. L'intervalle entre la présentation de 2 items était d'environ 1 seconde. Pour répondre, les sujets avaient à leur disposition un boîtier avec 2 boutons. Ils devaient effectuer une tâche de décision lexicale et devaient presser un bouton si l'item présenté était un mot et un autre si l'item vu n'était pas un mot de la langue française. La réponse 'mot' était associée à la main dominante du sujet. Les temps de réponses étaient mesurés par l'horloge de l'ordinateur, dotée d'une précision de +/- 1 ms. L'horloge était déclenchée par l'apparition de l'item à l'écran. Lorsque le sujet appuyait sur un bouton, l'horloge était stoppée, ce qui permettait la mesure des temps de réaction.

2.2 Sujets.

Vingt-trois sujets ont participé à cette expérience. Tous étaient de langue maternelle française, âgés de 20 à 25 ans.

2.3 Résultats

La première analyse vise à valider notre expérience et concerne la comparaison entre quatre catégories d'items: les Mots de Haute Fréquence (supérieure ou égale à 2 sur 1 million selon Lexique2 proposé par New, 2003), les Mots de Basse Fréquence (inférieure à 2), les Néologismes (Expérimentaux et de remplissage) et les Pseudomots.

Tout d'abord, il apparaît clairement (voir Figure 1) que les mots en général (MHF et MBF) sont traités plus rapidement que les pseudomots (728 ms vs. 792 ms). Ce résultat est classique, et s'explique vraisemblablement par le fait qu'avant de répondre 'non-mot', le système cognitif passe par des phases de vérification qui ne sont pas requises par la réponse 'mot'. Les néologismes quant à eux sont associés à des temps de décision bien plus importants mettant ainsi en évidence la difficulté de les classer.

De plus, nous observons l'effet classique de fréquence pour les mots : les mots de haute fréquence sont identifiés plus rapidement que les mots de basse fréquence. Cet effet de 63 ms est significatif par sujet et par item.

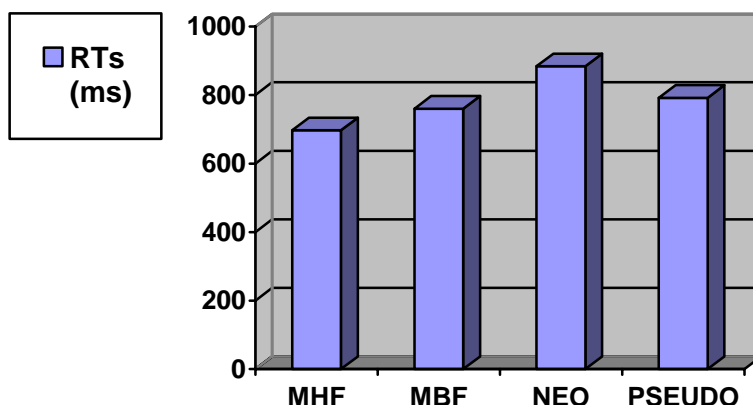


Figure 1 : Temps de réaction moyens pour les quatre types d'items : Mots de Hautes Fréquences (MHF), Mots de Basses Fréquences (MBF), néologismes (NEO), pseudo-mots (PSEUDO).

Au niveau des erreurs (voir Figure 2), on observe que les pseudo-mots et les mots de haute fréquence ont un taux comparable d'erreurs (4% et 5% d'erreurs), bien inférieur à celui observé pour les mots de basse fréquence (17%) et les néologismes (12%) mettant ainsi en évidence la facilité de la tâche pour les items non ambigus quant à leur lexicalité.

L'ensemble de ces résultats valide notre expérience.

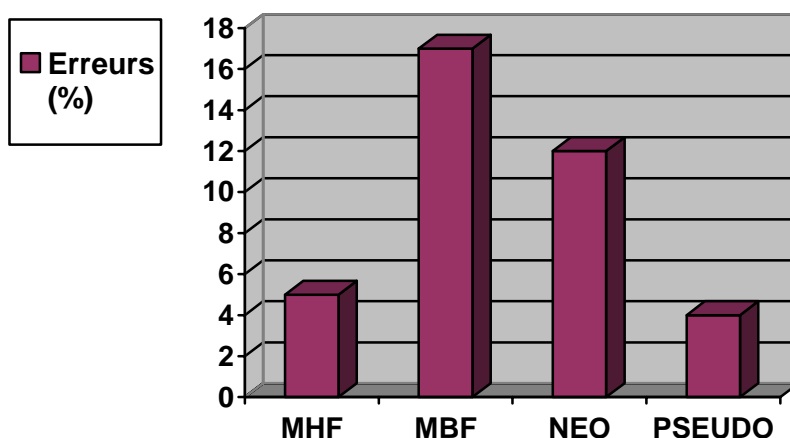


Figure 2 : Taux d'erreurs moyens pour les quatre types d'items : Mots de Hautes Fréquences (MHF), Mots de Basses Fréquences (MBF), néologismes (NEO), pseudo-mots (PSEUDO).

La suite de nos analyses va se concentrer sur les différentes catégories de néologismes expérimentaux.

Les temps de réponses (Figure 3) concernent les réponses correctes 'non-mots' et le taux d'erreur (Figure 4) correspond aux essais où les participants ont classé les néologismes expérimentaux comme un mot.

Dans un premier temps, il est à noter que l'analyse de l'effet de la fréquence cumulée de la famille morphologique de la base² sur les temps de réaction observés n'a pas donné de résultat significatif. Par ailleurs, une analyse de corrélations entre les temps de réaction et la fréquence cumulée de la famille morphologique de la base ou le log de celle-ci n'ont pas non plus donné lieu à des résultats statistiquement significatifs (respectivement $r_{65} = 0.07$ et $r_{65} = 0.18$).

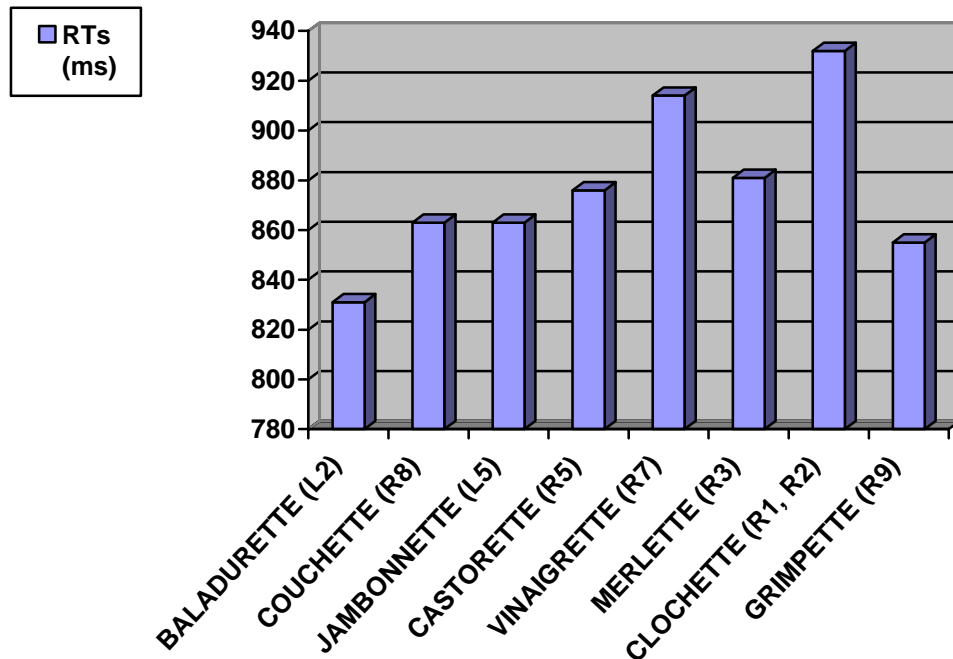


Figure 3 : Temps de réaction (RTs) moyens en millisecondes (ms) pour les huit types de néologismes : BALADURETTE-L2; COUCHETTE-R8; JAMBONNETTE-L5; CASTORETTE-R5; VINAIGRETTE-R7; MERLETTE-R3; GRIMPETTE-R9 et CLOCHETTE-R1,R2.

Les analyses statistiques spécifiques par sujet révèlent les faits suivants :

Selon les temps de réaction, le patron pour lequel la décision 'non-mots' est la plus facile à prendre est BALADURETTE-L2. Viennent ensuite [GRIMPETTE-R9, JAMBONNETTE-L5, COUCHETTE-R8, CASTORETTE-R5, MERLETTE-R3]³, et enfin [VINAIGRETTE-R7, CLOCHETTE-R1,R2].

Il semble donc que les patrons de construction de type VINAIGRETTE-R7 et CLOCHETTE-R1,R2 soient les plus disponibles, rendant ainsi la classification des néologismes appartenant à ces catégories comme non-mots lente.

² La fréquence cumulée de la famille morphologique de la base correspond à la somme de toutes les fréquences d'apparition de tous les lexèmes partageant une même base. Elle est calculée grâce à Lexique 2 (New & Pallier, 2003, <http://www.lexique.org>).

³ La mise entre crochets [...] signifie que les analyses statistiques inférentielles ne révèlent aucune différence entre les catégories incluses dans les crochets. Par exemple ici, les néologismes issus des patrons R9, L5, R8, R5 et R3 génèrent des temps de réaction comparables.

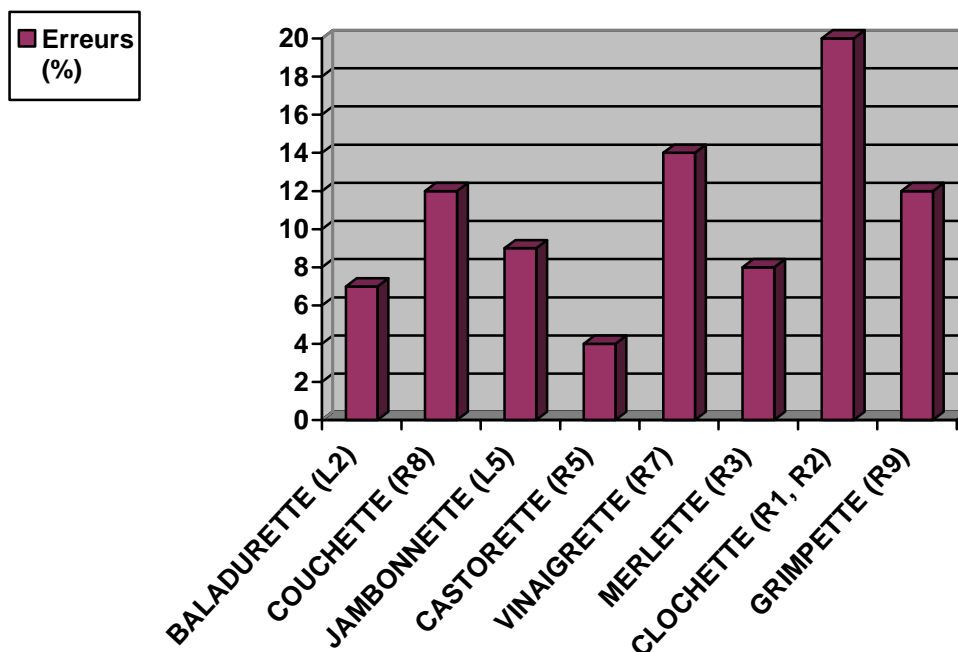


Figure 4 : Taux d'erreurs moyens pour les huit types de néologismes : BALADURETTE-L2; CASTORETTE -R5; MERLETTE -R3; JAMBONNETTE-L5; COUCHETTE -R8; VINAIGRETTE -R7; GRIMPETTE -R9 et CLOCHETTE -R1,R2.

Le pattern observé avec les temps de réaction est confirmé par le pattern d'erreur. Les patrons générant le moins d'erreurs sont [BALADURETTE-L2, CASTORETTE-R5]. Viennent ensuite [MERLETTE-R3, JAMBONNETTE-L5], [GRIMPETTE-R9, COUCHETTE-R8, VINAIGRETTE-R7], et enfin CLOCHETTE-R1,R2. Ce résultat suggère qu'il est facile de classer les néologismes issus des patrons [BALADURETTE-L2, CASTORETTE-R5] comme non-mots, alors que ceux issus du patron CLOCHETTE-R1,R2 sont difficilement classifiables (20% de réponses 'mot').

Globalement, les résultats mettent donc en évidence des patterns différents selon les patrons de construction des néologismes. Nous interprétons ces résultats en termes de disponibilité cognitive des différents patrons morphologiques. Pour résumer l'ensemble des résultats, il semble que CLOCHETTE-R1,R2 et VINAIGRETTE-R7 soient clairement des patrons productifs puisque les néologismes de ce type sont ceux qui prennent le plus de temps à être rejetés comme mots et ce sont également ceux pour lesquels il y a le plus d'erreurs (ils sont donc souvent classés comme mots). Les néologismes issus du patron BALADURETTE-L2 semblent eux être les plus faciles à classer comme non-mots : ils sont rejetés rapidement et donnent lieu à peu d'erreurs.

Afin de quantifier le rapport entre ces résultats et le calcul de productivité déjà effectué, nous avons mesuré les corrélations entre nos variables dépendantes (les temps de réactions et les taux d'erreur) et les mesures effectuées préalablement lors de notre analyse de corpus sur les dérivés transparents (le nombre d'hapax, le nombre d'occurrences et la mesure de productivité P – voir le Tableau 1). Nous observons une corrélation positive significative avec le nombre d'hapax (avec les temps de réaction, $r = 0.67$; $p = .05$; avec les erreurs, $r = 0.78$; $p < .05$), ce qui signifie que, plus il y a d'hapax recensés, plus les temps de réaction sont longs, donc le néologisme difficile à rejeter, et plus il y aura d'erreurs. On observe également une tendance de corrélation entre les erreurs et la mesure de productivité P effectuée sur le corpus pour ($r = -0.59$; $p = .08$) mais pas avec les temps de réaction ($r = -0.03$; n.s.). Il est à noter que le nombre

d'hapax et la mesure P ne sont pas corrélés ($r = -0.17$; n.s.). D'un point de vue descriptif, on observe des corrélations (bien que non significatives) entre nos résultats comportementaux et le nombre d'occurrences : plus il y a d'occurrences, plus les temps de décision pour juger les néologismes sont longs ($r = 0.32$; n.s.) et plus il y a d'erreurs ($r = 0.29$; n.s.). Cependant, il est important de souligner que ces corrélations sont à prendre avec précaution étant donné qu'elles sont effectuées uniquement sur huit types de néologismes.

Items / Patrons	Hapax	Occurrences	productivité
BALADURETTE (L2)	3	175	0,017
CASTORETTE (R5)	3	5	0,6
CLOCHETTE (R1, R2)	15	9071	0,0018
COUCHETTE (R8)	0	469	0
JAMBONNETTE (L5)	3	136	0,022
GRIMPETTE (R9)	10	573	0,017
MERLETTE (R3)	17	389	0,044
VINAIGRETTE (R7)	0	51	0

Tableau 1: Productivité des lexèmes construits transparents.

2.4 Discussion

Nos résultats présentent un pattern contrasté de temps d'identification pour les différents patrons de dérivation de *-ette*, cependant des incohérences apparaissent entre les résultats comportementaux et l'analyse de corpus effectuée, notamment pour le patron L2 qui avait été identifié comme productif sur la base de l'analyse de corpus et pour lequel il semble pourtant très facile de rejeter les néologismes construits. Une explication pourrait résider dans la bonne compréhension ou connaissance de l'item base. On peut effectivement se demander pour L2 si l'homme politique N-base a bien été identifié. Un autre problème réside dans le fait que certaines bases peuvent être ambiguës dans leur interprétation : par exemple, *moule* (*un moule* vs *une moule*) ; le patron en jeu serait alors différent. Il paraît dans ce cas souhaitable de connaître l'interprétation des individus. Dans une deuxième expérience, nous avons demandé aux participants de définir les néologismes utilisés dans l'Expérience 1.

3. Expérience 2

3.1 Méthode

Les néologismes utilisés dans l'Expérience 1 ont été répartis en listes de dix, où chacun des huit patrons était représenté au moins une fois, ceci afin que les sujets ne mettent pas au point des routines de réponses et réfléchissent pour chacun des items proposés. Les participants répondaient par écrit.

3.2 Sujets

Soixante sujets, âgés de 20 à 25 ans, ont participé à cette expérience. Tous avaient le français pour langue maternelle.

3.3 Résultats

Chaque néologisme a été défini par au moins cinq individus.

Pour le patron CLOCHETTE-R1,R2, on n'observe qu'une seule non-réponse. La définition 'petit N-base' est très largement présente (98% des réponses). Seul le néologisme *grangette* génère une réponse différente, i.e. « type de croquette ».

Pour le patron MERLETTE-R3, on observe peu de non-réponses (4%). 40% des définitions proposées évoquent l'aspect 'femelle' et 40% l'aspect 'petit'. Il est à noter que les notions de 'jeune' et 'bébé' sont également utilisées (27%). On observe dans cette catégorie davantage de diversité dans les réponses obtenues que dans la catégorie précédente : pour *lézardette* par exemple, on obtient « petite femelle lézard » mais aussi « tissu en forme de lézard », pour *tigrette*, « petit tigre femelle » mais également « personne qui grogne », pour *hamsterette*, « petite femelle hamster » mais aussi « joli hamster » et « cage pour hamster », pour *fouinette*, « petite fouine », « personne malicieuse qui fouine » et « petit judas de porte ». Il est à noter que, pour *tapirette* la base ne semble pas avoir été identifiée et les participants ont préféré utiliser *tapis*, i.e. « tapis miniature », « serpillière », « tapis volant ». Un élément intéressant concerne les réponses observées pour *tigrette* qui a été construit sans tenir compte du possible phénomène de blocage dû à l'existence de *tigresse* qui pourrait avoir bloqué l'interprétation 'femelle du tigre'. Il est effectivement observable que la plupart des définitions données sont « la petite tigresse » et qu'une seule réponse propose la synonymie « femelle du tigre ».

Pour la catégorie CASTORETTE-R5, L1f, on observe quelques non-réponses (13%). La notion de 'petit' est présente dans 40% des cas comme celle de 'femelle'. La notion de 'bébé' est évoquée dans uniquement 2% des cas. Cette catégorie révèle que certains N-base utilisés pour construire les néologismes ne sont pas connus par nos participants. Par exemple pour *vairrette*, on obtient comme définitions : « petite lucarne », « petite verrerie », « petite vérole » et une non-réponse. De même pour *zibelinette* : « petite zibeline », « style de trompette », « petite zigounette », « petit briquet » et une non-réponse. De plus, même lorsque la base est identifiée, les réponses sont variées comme pour *herminette*, « petite femelle hermine », « bébé de l'hermine », « femelle de l'hermine », « gant en fourrure » et « héroïne de BD ». On observe parfois aussi l'utilisation d'un sens métaphorique de la base, *renardette*, « fille intelligente ». Force est de constater qu'il existe une grande similitude entre les réponses obtenues pour le patron MERLETTE-R3 et celles obtenues pour le patron CASTORETTE-R5,L1f, ce qui s'explique facilement par le fait qu'a priori il n'est pas possible de savoir si le N-base désigne l'animal ou sa fourrure.

Pour la catégorie VINAIGRETTE-R7, on observe peu de non-réponses (2%). 'Petit' est cité dans 46% des cas, et 'femelle' dans 2%. Cette catégorie se caractérise par une grande hétérogénéité des définitions données. Les définitions données à *basiliquette* illustrent bien ce constat : « femelle du basilique⁴ », « sauce au basilic », « petite brindille de basilic », « plat frais à base de basilic », « soupe au basilic ». De même pour *persillette* sont proposés « une petite boîte pour mettre du persil », « instrument pour broyer le persil », « petite pincée de persil », « persillade sans ail ». A nouveau, on observe différents sens attribués au N-base: *aillette*, « petite aile », « sauce à l'ail »; ou *poivrete*, « sauce au poivre » mais aussi « femme alcoolique », ainsi que la non compréhension du N-base : pour *cardamomette*, « thermomètre

⁴ L'orthographe utilisée dans les définitions est celle majoritairement utilisée par les participants.

de nouvelle génération », « base de TNT ». 20% des définitions tournent autour de substances obtenues à partir de N-base, mais la notion de non majoritaire que nous avons identifiée n'est absolument pas exprimée.

Dans la condition COUCHETTE-R8, il y a 3% de non-réponses, 46% des réponses utilisant 'petit' et 3% de réponses faisant référence à la notion de 'femelle'. On observe à nouveau dans cette catégorie une grande difficulté à donner une définition, malgré la connaissance de la base. Un exemple clair est donné par *mangette* puisqu'ici, on ne peut suspecter la non connaissance de la base. Or les définitions données sont : « boisson à base de mangue », « petite mangue pas encore mûre », « petite assiette creuse », « manger un peu » ainsi qu'une non-réponse. De plus, aucun participant ne propose clairement une définition de lieu, que la base soit identifiée ou non. Par exemple, pour *dormette* où la base est bien identifiée, on obtient : « un objet permettant de s'endormir, comme un doudou », « petite sieste » et « sieste au soleil » On peut également observer que parfois les participants préfèrent modifier la base probablement pour retrouver un patron de dérivation plus disponible comme pour *séchette*, « femelle de la sèche ».

Pour la condition GRIMPETTE-R9, on observe davantage de non-réponses (20%) que dans les précédentes conditions et peu de réponses 'femelle' (3%). 53% des réponses évoquent la notion de petitesse. En fait, les réponses semblent bien refléter la notion de courte durée comme pour *téléphonette* « courte conversation téléphonique ». Mais il est vrai que cette analyse dépend souvent fortement de l'acceptation que l'on attribue à 'petit'. Par exemple « petite glissade », définition souvent proposée pour *glissette*, peut vouloir dire 'de courte durée' mais aussi de façon liée 'sur une courte distance'. On observe parfois à nouveau une hétérogénéité des réponses : *renversette*, « jeu pour enfant du genre de balançoire », « défaite inattendue », « révolution dans un petit groupe », « rocking chair » et « petite pente ».

Pour la catégorie BALADURETTE-L2, on observe plus de non-réponses (23%) que dans les autres conditions, et quelques « petit » (10%). La majorité (51%) évoque une féminisation sous forme de « petite fille de », « fan de », « femme de », « maîtresse de », « danseuse de », « majorette » et « groupie », ce qui signifie que le N-base a bien été identifié comme un nom propre. Cependant, il apparaît également que certains noms d'hommes politiques ne sont pas identifiés et/ou connus : *devaquette* est défini par « nom de famille », « blouson pour manchot », « siège en acajou utilisé à la Renaissance » ainsi que deux non-réponses. De même *toubonnette* reçoit les définitions « chose tout bien », « femme de Toubon », « gentille femme », « tourte aux courgettes ». Aucune mention n'est faite de mesures administratives.

Pour la condition JAMBONNETTE-L5, on observe quelques non-réponses (11%) et pas de féminisation. A nouveau, 63% des réponses font référence à la petitesse. Il est à noter que 37% des réponses définissent un type d'alimentation. Il est également remarquable que certaines réponses réfèrent à un contenant : « petit verre à bière », « petit récipient pour mettre de l'huile », et également à des outils, « instrument permettant d'étaler de l'huile dans les plats ».

3.4 Discussion

Il ressort de cette analyse que les définitions fournies par les sujets sont très diverses et ne correspondent pas forcément aux patrons utilisés pour leurs créations. Pour partie, ces résultats peuvent être expliqués par des faiblesses du matériel. Par exemple, certaines bases

utilisées sont ambiguës, c'est le cas pour *basiliquette*, par exemple, qui aurait pu être construit sur *basilic* 1- 'plante aromatique', *basilic* 2- 'serpent venimeux', *basilique* 1- 'église', *basilique* 2- 'veine'. Cependant, les réponses observées ne soulignent pas particulièrement ce problème de polysémie des bases choisies. En revanche, une caractéristique qui semble être reflétée par nos résultats est qu'un mot-forme dérivé d'une seule et même base peut représenter plusieurs patrons (voir Plénat, 2005). C'est le cas par exemple des formes [patronyme + *ette*] qui peuvent effectivement nommer des mesures administratives (*balladurette*), mais également des assistantes féminines (*juppettes*), des artefacts (*chiraquette*) et des sexuisemblanciers ('épouse de'). Cela dit, c'est aussi le but de nos expérimentations de voir dans ces cas quel type d'interprétations est choisi, et c'est même ce qui justifie notre deuxième expérience mais module effectivement les résultats de la première. Un autre point de discussion concerne la vraisemblance phonologique des constructions. Certains néologismes étaient clairement moins vraisemblables que d'autres, comme par exemple *serpentette*, qui enfreint une contrainte dissimilative. Il se peut que ce facteur ait joué sur les résultats de la première expérience : le sentiment de bizarrerie phonologique entraînant un rejet plus rapide de l'item comme mot.

Malgré ces faiblesses, l'approche expérimentale proposée fournit des informations intéressantes sur l'interprétation de lexèmes complexes inconnus. Par exemple, on voit une prédominance de la notion de petitesse dans toutes les catégories, avec toute l'ambiguïté que peut porter ce terme. La notion de 'féminin' revient également dès que le référent du N-base le permet, c'est-à-dire dès qu'il s'agit d'un animé. On note également que les notions utilisées dans nos descriptions des types de dérivés sont quasiment toutes présentes dans au moins une définition, sauf pour les mesures administratives de L2. Le cas de L2 semble très particulier car, bien qu'ayant été identifié comme très productif dans l'analyse de corpus, il apparaît peu disponible dans la validation expérimentale. Une explication possible réside dans le fait que la productivité de ce patron est exclusivement restreinte au milieu journalistique (le corpus utilisé pour l'analyse était le quotidien *Libération*). Or les étudiants testés suivaient essentiellement un cursus de psychologie, connu pour une certaine non-politisation. Il est donc envisageable que les jeunes participants à nos expériences n'aient jamais été exposés à des lexèmes issus de ce patron. Des études supplémentaires sont nécessaires afin d'éclaircir ce point. Il faut également garder à l'esprit que ce qui est formulé consciemment par le sujet n'est pas forcément le reflet exact des processus cognitifs sous-jacents à l'identification des lexèmes morphologiquement complexes. L'idéal serait de partir des définitions obtenues pour créer du matériel mieux contrôlé afin de refaire une expérience en temps réel utilisant les temps de réaction. Le travail expérimental présenté dans cet article constitue un premier essai d'application de méthodes expérimentales à des questions linguistiques précises sur la productivité morphologique. Bien que loin d'être parfait, notamment dans la construction du matériel linguistique, ce travail présente une approche pluridisciplinaire de toute évidence fructueuse et informative. Des recherches ultérieures viseront à améliorer et à raffiner ce type d'études.

4. Conclusion

Dans cet article, nous avons exposé comment, à partir d'une description linguistique de la productivité de la suffixation en -ET, nous pouvions tester la validité cognitive de cette analyse en proposant une expérimentation utilisant des paradigmes en temps réel comme la décision lexicale mais également des tâches plus métalinguistiques comme trouver une définition au lexème. Un va et vient entre ces différentes approches peut permettre de mieux circonscrire l'utilisation et la compréhension des règles de construction des lexèmes d'une langue.

5. Bibliographie

- Aliquot-Suengas, S. (1997), "Les suffixes -ade et -é(e): une histoire de famille", in B. Fradin & J.-M. Marandin (eds), *Mot et grammaires*, Paris: Didier Erudition, pp. 49-87.
- Baayen, H. R., & Renouf, A. (1996), "Chronicling the Times: Productive lexical innovations in an English newspaper", *Language* 72, pp. 69-96.
- Baayen, H., & Lieber, R. (1991), "Productivity and English derivation: a corpus-based study", *Linguistics* 29(5), pp. 801-843.
- Corbin, D. (1987), *Morphologie dérivationnelle et structuration du lexique*, 2 vol., Tübingen, Max Niemeyer Verlag ; rééd., Villeneuve d'Ascq, Presses Universitaires de Lille, 1991.
- Dal, G. (1997), *Grammaire du suffixe -et(te)*, Paris: Didier Erudition.
- Delhay, C. (1996), *Il était un "petit X". Pour une approche nouvelle de la catégorisation dite diminutive*, Collection Sciences du langage, Paris: Larousse.
- Dell, G. S. (1986), "A spreading-activation theory of retrieval in sentence production", *Psychological Review* 93, pp. 283-321.
- Fradin, Bernard. 2003. « Le traitement de la suffixation en **-ET** ». *Langages*:51-77.
- Fradin, B., Hathout, N. & Meunier, F. (2003). "La suffixation en -ET et la question de la productivité", in G. Dal (Ed.), *Langue Française* 140.
- Frauenfelder, U. H., & Schreuder, R. (1992). "Constraining psycholinguistics models of morphological processing and representation: the role of productivity", in G. Booij & J. van Marle (Eds.), *Yearbook of morphology 1991* (pp. 165-183). Netherlands: Kluwer Academic Publishers.
- Manelis, L., and Tharp, D. A. (1977), "The processing of affixed words", *Memory and Cognition* 5, pp. 690-695.
- Plag, Ingo, Christiane Dalton-Puffer, & Harald Baayen. 1999. "Morphological Productivity across speech and writing". *English Language and Linguistics* 3:209-228.
- Plénat, M. (2005). "Brèves remarques sur les déverbaux en -ette". In F. Lambert & H. Nølke (éds), *La syntaxe au coeur de la grammaire. Recueil offert en hommage pour le 60e anniversaire de Claude Muller*. Rennes, Presses Universitaires de Rennes, pp. 245-258.
- Schultink, H. (1962), "De morfologische valentie van het ongelede adjectief, proefschrift Leiden; ongewijzigde herdruk uitgegeven door HES Publishers, Utrecht 1980.
- Stemberger, J. P. (1985). "An interactive activation model of language production", in A. W. Ellis (Ed.), *Progress in the psychology of language* Vol. 1, Hillsdale, NJ: Erlbaum, pp. 143-186.
- Taft, M. & Forster, K. I. (1975), "Lexical storage and retrieval of prefixed words", *Journal of Verbal Learning and Verbal Behavior* 18, pp. 638-647.

Annexe : Néologismes utilisés dans les Expériences 1 et 2

*R1 et R2 :
clochette*

bouteillette
cavette
grangette
blousette
tranchette
ampoulette
phrasette
armoïrette
épongette
biblette

*R5 et L1f :
castorette*

herminette
renardette
vairette
lézardette
cashemirette
chinchillette
panthérette
ragondinette
zibelinette
lapinette
serpentette

R8 : couchette

dormette
visionnette
étalette
baignette
mangette
séchettes
joggette

L2 : baladurette

giscardette
devaquette
toubonnette
fabiusette
gaudinnette
sarkosette
chiraquette
jauressette

R3 : merlette

lézardette
hérissonnette
tigrette
tapirette
fouinette
dromadaïrette
girafette
hamsterette

R7 : vinaigrette

basiliquette
persillette
fraisette
cerisette
cardamomette
poivrete
aillette
goyavette
papayette

R9 : grimpette

glissette
souplette
téléphonette
observette
arrosette
renversette

L5 : jambonnette

moulette
semoulette
oeufette
bièrette
oignonnette
huilette