

PRODUCTIVITÉ ET TAILLE DE LA FAMILLE MORPHOLOGIQUE: LE CAS DES NOMS COMPOSÉS FRANÇAIS N-de-N

Elena Voskovskaia
Université de Toronto

Résumé Des travaux récents dans la morphologie dérivationnelle ont démontré un nouvel intérêt pour le concept de la productivité morphologique et pour ses outils de mesure en utilisant une méthode quantitative (Baayen 1992, 1993; Baayen et Lieber 1991; Baayen et Renouf 1996) et qualitative (Bauer 2001; Dal 2003). Cependant, la productivité des mots composés a été très peu explorée, particulièrement en français, à part le travail de Krott, Schreider et Baayen (1999). Récemment, une nouvelle approche à la mesure de la productivité morphologique en terme de la taille de famille morphologique a été proposée (de Jong, Schreuder et Baayen, 2000; del Prado et al. 2004). Ce papier étudie une corrélation possible entre la productivité morphologique et la taille de la famille morphologique des composés français *N-de-N*. Le résultat de la recherche indique qu'il existe une corrélation positive entre la productivité et la taille catégorielle de la famille morphologique dans les composés N-de-N où la croissance de la famille morphologique augmente le niveau de la productivité de ces composés.

Abstract Recent works in derivational morphology have shown a new interest for the concept of morphological productivity and for its various measuring tools, both quantitative (Baayen 1992, 1993; Baayen & Lieber 1991; Baayen & Renouf 1996) and qualitative (Bauer 2001; Dal 2003). However, the productivity of compound words has been very little explored, especially in French, except for the work of Krott, Schreider & Baayen (1999). Recent studies, have proposed a new approach to morphological productivity in term of family size (De Jong, Schreuder, & Baayen, 2000, del Prado M., Bertram, et al). This paper considers a possible correlation between morphological productivity and family size and based on the results of a quantitative study of French compounds *N-de-N*. The result of study indicates that there is a positive correlation between the productivity and family size of compounds N-de-N. The growth of the morphological family increases the level of the productivity of these compounds.

Mots-clefs : Morphologie, la productivité morphologique, corpus linguistique, composés français

Keywords: Morphology, morphological productivity, linguistic corpus, French compounds

1. Instruction

La notion de productivité morphologique peut être abordée sous un angle qualitatif ou quantitatif. D'un point de vue qualitatif, on définit la productivité comme la capacité des règles à former de nouveaux lexèmes de façon non intentionnelle (Dal 2003: 5) à partir de la disponibilité du processus morphologique (Bauer 2001:49). D'un point de vue quantitatif, plusieurs méthodes à base de corpus ont été proposées pour mesurer des aspects différents de la productivité (Baayen et Lieber 1991; Baayen 1992, 1993; Baayen et Renouf 1996; Krott et al.1999). Toutefois, exception faite des travaux de Krott, Schreider et Baayen (1999) et Fernández-Domínguez et al. (2007) les études portant sur la mesure de la productivité n'ont pratiquement pas traité des mots composés.

L'ouvrage de Krott, Schreider & Baayen (1999) révèle un phénomène intéressant, soit que le degré de la productivité des composés diffère substantiellement de la productivité de ses constituants. Il existe en fait une corrélation entre la structure morphologique des mots-bases et la productivité des mots composés comme structure figée: la fréquence et la longueur des constituants employés seuls déterminent leur probabilité à l'intérieur des composés. Les mots qui sont plus fréquents et plus courts ont plus de chances de devenir les constituants des composés. Cela soulève un paradoxe de productivité morphologique propre aux mots composés (Krott et al. 1999 : 914): les catégories de mots qui possèdent un degré de productivité assez bas sont relativement plus productives comme constituants des composés, tandis que les catégories de mots qui sont plus productives en général sont moins productives comme constituants des composés.

Les mesures de calcul de la productivité les plus utilisées actuellement sont présentées par Baayen (1991, 1996, 1999). Baayen propose deux mesures de calcul de la productivité morphologique : *la productivité au sens strict* et *la productivité globale*; les deux mesures se calculent relativement à un corpus donné. La *productivité globale* (global productivity), notée P^* , exprime la probabilité d'apparition d'un mot nouveau dans un procédé donné (Hay & Baayen 2002).¹ La mesure la plus utilisée dans les recherches sur la productivité morphologique est actuellement celle 'au sens strict' (Baayen et Lieber :1991; Baayen 1992). La *productivité au sens strict*, notée P , est représentée comme mesure quantitative qui se base sur la notion d'hapax legomenon.

À côté d'une analyse quantitative traditionnelle de la productivité morphologique, une autre approche de la mesure de la productivité est régénérée récemment dans les milieux linguistiques – celle de la corrélation entre la productivité et la taille de la famille morphologique. Prado et al. (2005:496) montrent que la taille de la famille morphologique est fortement corrélée avec la fréquence du mot, c'est-à-dire, si un mot est fréquent, la taille de sa famille morphologique a tendance à être plus grande. La taille de la famille catégorielle

¹ Dans ses premiers travaux, H. Baayen représente la productivité globale comme un moyen d'incorporer la rentabilité du procédé étudié dans la mesure de la productivité (deux procédés A et B peuvent en effet avoir la même valeur P alors que l'un forme dix fois plus de mots que l'autre).

est une mesure à laquelle une catégorie morphologique au complet contribue à la connectivité morphologique dans le lexique mental.

En analysant la corrélation entre la productivité des affixes et la taille de la famille morphologique en anglais, Baayen et Hay (2002) montrent que l'augmentation de la taille d'une famille catégorielle cause la croissance de la productivité morphologique. Ils définissent la taille catégorielle de la famille morphologique comme suit: « ...the 'categorical family size' of a morphological category be the sum of the (logarithms of the) family sizes of the base words of the members of this category" (Baayen et Hay 2002:1). Par exemple, la taille de la famille catégorielle de l'affixe *un-* dans les mots *unbound*, *unsafe*, *unbrave*, *unwise* est la somme $\log(F1) + \log(F2) + \log(F3) + \log(F4)$ etc. où F1, F2, F3, sont les tailles de la famille des mots-bases *bound*, *safe*, *brave*, *wise*, etc. La corrélation entre la taille catégorielle de la famille morphologique et le niveau de la productivité était notée pour le processus dérivationnel en anglais.

La mesure de la productivité morphologique, très peu étudiée pour les noms composés, soulève une question importante: « Est-ce qu'il existe une corrélation entre la productivité et la taille de la famille morphologique des composés ? » Notre hypothèse est la suivante: Il existe une corrélation directe entre la productivité et la taille de la famille morphologique des composés *N-de-N*: la croissance de la taille catégorielle de la famille morphologique augmente le niveau de leur productivité.

En utilisant la mesure *P* de la productivité *au sens strict* de Baayen (1991, 1992) et le modèle de la taille catégorielle de la famille morphologique de Baayen et Hay (2002), ce papier examine la productivité des noms composés français *N-de-N* et vise à explorer la corrélation potentielle entre la productivité des noms composés en français et la taille catégorielle de leur famille morphologique.

La recherche est basée sur le corpus textuel *Frantext* formé plus de 4000 textes en utilisant 49 composés *N-de-N* qui sont extraits et étudiés à partir du Dictionnaire de Littré (1877-1878).

Dans la suite de cet article, nous présentons les méthodes de calcul de la productivité utilisées dans la recherche (sec. 2), la description du corpus et la préparation du matériel étudié (sec. 3). Nous discutons les résultats obtenus sur la productivité des composés *N-de-N* (sec. 4) et la corrélation possible entre la productivité et la taille catégorielle de leur famille morphologique (sec.5). Nous terminons avec une conclusion et des perspectives (sec. 6).

2. Méthodes du calcul de la productivité

Dans notre travail, deux mesures sont utilisées pour évaluer la productivité des composés *N-de-N*: la mesure basée sur la notion d'hapax legomenon (Baayen et Lieber 1991: 809; Baayen 1992: 115-16; Hay et Baayen 2003: 101) et la mesure de la taille de la famille d'une catégorie morphologique de Baayen et Hay (2002).

L'évaluation basée sur les hapax est une mesure relative qui doit comparer la productivité des procédés morphologiques à l'intérieur d'un même corpus ou entre des corpus différents (Baayen & Renouf 1996). La méthode

quantitative basée sur des hapax mesure la productivité morphologique selon la formule suivante :

$$(1) \quad P = n_1 / N$$

où 'P' représente la productivité dans le sens étroit (le taux de croissance du vocabulaire), n_1 représente le nombre d'hapax legomenon (les formes qui apparaissent une fois seulement dans un grand corpus), et N représente le nombre total d'occurrences observées dans le corpus contenant le type dérivationnel analysé.

Selon cette mesure de productivité, une catégorie avec un grand nombre de mots de haute fréquence aura une grande valeur de N et, en conséquence, un plus bas degré de productivité. Ce calcul évalue la productivité du gabarit (du type) des noms composés comme une structure figée en démontrant le taux de croissance du vocabulaire. Pour que les hapax représentent de vrais néologismes, il faut que la recherche de P soit faite sur un très grand corpus (des dizaines de millions de mots).

La deuxième mesure impliquée dans l'évaluation de la productivité morphologique des composés N-de-N est celle de la taille de la famille d'une catégorie morphologique proposée par Baayen et Hay (2002). Dans leur travail sur la productivité des affixes et des bases, Baayen et Hay (2002 :1) montrent qu'il existe une forte corrélation entre le degré de la productivité de la base et la taille de la famille morphologique. Les mots comme *man* ou *coin* sont présentés comme constituants (des bases) dans plusieurs mots complexes; ils possèdent un haut degré de productivité et leurs familles morphologiques sont assez grandes. Cet effet lié à la taille de la famille morphologique a été aussi observé en néerlandais, allemand, anglais, hébreu et finlandais.

La mesure de la taille de la famille d'une catégorie morphologique se calcule en deux étapes :

Premièrement, on mesure la taille de la famille morphologique de la forme recensée (par exemple, la famille du mot *chant*, inclut tous les dérivés et les composés formés à partir de cette base, sauf la base elle-même : *chanteur*, *chanter*, *chanterelle*, etc. Selon la définition de Schreuder et Baayen (1997 : 118), la famille morphologique inclut tous les mots qui sont formés avec le mot de base via la dérivation ou la composition, mais pas la flexion. Le terme *famille morphologique* est utilisé par Schreuder et Baayen (1997 :121) pour dénoter une série de mots dérivés d'un radical *table* donné à travers la composition (*tablespoon*, *timetable*) ou la dérivation (*tablet*, *tabular*).

Deuxièmement, on évalue la *taille catégorielle de la famille* morphologique qui est la somme des logarithmes des tailles de famille des mots de base, membres de cette catégorie. Pour mesurer la *taille catégorielle de la famille* morphologique des composés N-de-N, on utilise une formule logarithmique suivante proposée par Baayen et Hay (2002 :1) :

$$(2) \quad \log (F_1) + \log (F_2) + \log (F_3) \text{ etc.}$$

où F1, F2, F3, sont les tailles de la famille morphologique des mots-bases de ces composés.²

La taille de la famille catégorielle est une mesure à laquelle une catégorie morphologique au complet contribue à la connectivité morphologique dans le lexique mental.

Dans notre travail, la mesure de la taille catégorielle de la famille morphologique calcule la fréquence des constituants du composé aussi bien que la fréquence des membres de leurs familles morphologiques.

3. Corpus et préparation du matériel

Plusieurs mesures de la productivité nécessitent l'extraction des données à partir de grands corpus. Les hapax qui y apparaissent sont de bons marqueurs de la productivité du processus étudié (c'est-à-dire de leur aptitude à former des mots nouveaux) : plus grand est le nombre d'hapax – plus productif est le processus morphologique étudié. C'est pourquoi la taille du corpus est extrêmement importante : un corpus trop limité risque de présenter à tort des formes uniques comme des néologismes.

Vu la nécessité d'effectuer une recherche sur un grand corpus représentatif, deux sources de données sont utilisées dans notre recherche.

Le *Dictionnaire de la langue française de Littré* (édition de 1877-1878) est utilisé pour former la liste de noms composés N-de-N et établir notre propre corpus de ce type de composés. Nous avons recours à cette base de données puisqu'elle est très représentative. Avec ses 78 423 entrées qui englobent une période allant du XVI^e au début du XIX^e siècle, ce dictionnaire représente tous les termes usuels des sciences, des arts, des métiers et de la vie pratique de son époque aussi bien que des citations des ouvrages littéraires (Voltaire, Corneille, Montaigne, La Fontaine, etc.). C'est un dictionnaire normatif qui reflète un état de la langue française classique et le bon usage littéraire entre le XVII^e et le XIX^e siècle et qui demeure une étape importante dans le développement de la langue française caractérisée par Dantier (2003) comme « la langue entre la lexicographie et la linguistique ». Nous avons utilisé l'édition 1877-1878 de ce dictionnaire disponible en CD. Pour extraire des formes *N-de-N*, le dépouillement manuel du dictionnaire était appliqué pour créer la liste des noms composés à trait d'union. La liste de 49 composés *N-de-N* (*bec-de-cane*, *cul-de-four*, *patte-d'oie*, etc.) formé à partir du Dictionnaire de Littré (1877-1878) sera analysée plus tard dans le corpus *Frantex*.

Après avoir établi la liste et la typologie des noms composés qui seront mesurés, la fréquence d'occurrences de ces formes a été validée dans un corpus représentatif de grande taille. Nous avons recours au corpus textuel *Frantex*

² En suivant Schreuder et Baayen (1997), Moscoso del Prado et al. suggèrent aussi que l'effet de la taille de la famille est logarithmique par la nature : « Like the word-frequency effect, the family-size effect is logarithmic in nature. Robust effects are typically observed in the range of 0–40 family members, after which there is generally a floor effect ». (Moscoso del Prado et al. 2004 :1272)

disponible en ligne. La base de données Frantex³ est un large corpus informatisé composé de textes en français, créé par le laboratoire ATILF. Cette base de données avec 4000 textes couvre des domaines tels que les sciences, les arts, la littérature, les techniques, la politique, la philosophie, etc. et elle a une taille suffisante pour appliquer une série de mesures quantitatives. Ce corpus est divisé en quatre périodes importantes dans l'histoire de la langue française : 1606-1694 (17.3 millions de mots) ; 1695-1798 (34.4 millions de mots) ; 1799-1872 (41 millions de mots); 1873-1920 (28 millions de mots).

La version non catégorisée de Frantext a été utilisée dans notre recherche parce que la version catégorisée est limitée aux textes postérieurs à 1829. La version non catégorisée de Frantext nous permet de faire toutes les analyses statistiques nécessaires (le nombre d'occurrences, de néologismes, de membres des familles morphologiques, etc.). Les statistiques sur la fréquence d'occurrences des composés N-de-N dans *Frantext* paraissent au Tableau 1 :

Tableau 1 Fréquence des composés N-de-N dans le corpus *Frantext*

	1606-1694	1695-1798	1799-1872	1873-1920	Total
Composés	113	173	204	277	767
Œuvres	141	183	283	64	671
Mots	4,884,725	12,524,257	20,608,307	5,208,641	43,225,930

Les résultats du Tableau 1 montrent que la fréquence des composés N-de-N est plus élevée pendant la troisième et quatrième période (204 et 277 occurrences respectivement).

4. Productivité des composés N-de-N: résultats et discussion

Au cours de nos recherches sur les composés *N-de-N* dans le corpus Frantex , 23 hapax legomenon (les formes qui apparaissent une fois seulement dans un grand corpus) ont été repérés. En appliquant la mesure de la productivité de Baayen (1991:809), on a calculé la productivité générale des noms composés *N-de-N* en français du 17^e au début du 20^e siècle comme suit :

$$(3) P = \frac{23}{767} = 0.030$$

³ La base de donnée de Frantext (<http://humanities.uchicago.edu/orgs/ARTFL/>) fut élaborée en 1960 à partir d'un premier ensemble de textes pour apporter une documentation d'exemples au *Trésor de la langue française*. Elle comportait moins de mille œuvres des XIX^e et XX^e siècles sélectionnées à partir des bibliographies de manuels d'histoire littéraire. Cet ensemble s'est rapidement enrichi vers 1970 de textes scientifiques et techniques nécessaires à la rédaction du dictionnaire. Par la suite, le fonds s'est étendu de façon à couvrir les trois siècles précédents ainsi que la période suivante. Le corpus continue de s'étendre progressivement.

où 23 = le nombre total d'hapax dans le corpus
767 = le nombre total d'occurrences des composés *N-de-N* dans le corpus

Les changements dans le taux de la productivité morphologique des composés *N-de-N* ont démontré des résultats suivants d'une période à l'autre :

Tableau 2 Productivité des composés *N-de-N* dans le corpus *Frantext*

Période	n ₁	N	P
1606-1694	5	113	0.044
1695-1798	3	173	0.017
1799-1872	6	204	0.029
1873-1920	9	277	0.032
Total :	23	767	0.122

où n₁ = le nombre total d'hapax
N = le nombre total d'occurrences des composés *N-de-N*
P = la productivité morphologique des composés *N-de-N*

L'analyse du Tableau 2 permet de dégager les conclusions suivantes :

1. Les composés *N-de-N* étaient les plus productifs pendant la période 1606-1694 (P=0.044).
2. La productivité des composés *N-de-N* était similaire pendant les périodes 1799-1872 et 1873-1920 (P=0.029 vs P=0.032).
3. La productivité morphologique de ce type de composés s'est abaissée 2.5 fois au cours des années 1695-1798 (P=0.012) par rapport à la première période étudiée (1606-1694).

5. Productivité et taille catégorielle de la famille morphologique des composés N-de-N

La mesure de la productivité morphologique, très peu étudiée pour les noms composés français, soulève une question importante: « Est-ce qu'il existe une corrélation potentielle entre la productivité et la taille de la famille morphologique des noms composés ? »

La corrélation entre la taille catégorielle de la famille morphologique et le niveau de productivité a été notée pour les processus d'affixation en anglais. Baayen et Hay (2002) ont découvert que l'augmentation de la taille d'une famille catégorielle cause la croissance de la productivité morphologique dans les mots dérivés (Baayen et Hay 2002:1).

Pour mesurer la taille de la famille morphologique des composés *N-de-N*, nous avons dégagé tous les mots appartenant à la même famille morphologique que chacun des constituants du mot composé en question. Par exemple, pour le composé *belle-de-jour*, nous avons cherché dans *Frantext* des mots comme *beau*, *beauté*, *bellâtre*, *bonjour*, *toujours*, *séjourner*.

La taille catégorielle de la famille morphologique des composés *N-de-N* a été évaluée en deux étapes:

1) Premièrement, la taille de la famille morphologique des composés *N-de-N* a été calculée par quatre périodes comme F1=63; F2=70; F3=90 et F4=102 respectivement. Selon la définition de Schreuder et Baayen (1997 :118), la famille morphologique inclut tous les mots qui sont formés avec le mot de base via la dérivation ou la composition, mais pas la flexion, exclut le mot de base.

La distribution des membres des familles morphologiques des composés *N-de-N* par période est présentée dans le Tableau 3 :

Tableau 3. Les composés N-de-N dans le corpus *Frantext*

	1606-1694	1695-1798	1799-1872	1873-1920	Total :
Composés	113	173	204	277	767
Œuvres	141	183	283	64	671
Mots	4,884,725	12,524,257	20,608,307	5,208,641	43,225,930

2) Deuxièmement, la formule logarithmique de Baayen et Hay (2002 :1) a été appliquée. La taille catégorielle de la famille d'une catégorie morphologique est calculée comme la somme des logarithmes des tailles de famille des mots de base. Par exemple, pour la première période la taille catégorielle du composé *N-de-N* avec 12 types différents était calculée comme suit :

$$(4) \quad \log(4)+\log(3)+\log(6)+\log(6)\dots=8.17$$

La corrélation entre le taux de la productivité et la taille catégorielle de la famille morphologique des composés *N-de-N* est récapitulée dans le Tableau 4 :

Tableau 4 Productivité et taille catégorielle de la famille morphologique des composés *N-de-N* par période

mesure	1606-1694	1695-1798	1799-1872	1873-1920	Total
N	113	173	204	277	767
F	63	70	90	102	325
T	8.17	8.42	11.66	13.56	41.81
P	0.044	0.017	0.029	0.032	0.122

où N = le nombre total d'occurrences des composés *N-de-N*
 F = la taille de la famille morphologique des composés *N-de-N*
 T = la taille catégorielle de la famille morphologique des composés *N-de-N*

P = la productivité morphologique des composés *N-de-N*

Le Tableau 4 représente la croissance du vocabulaire des composés N-de-N. On observe, en particulier, la corrélation directe entre la productivité et la taille catégorielle de la famille morphologique de ces composés à travers trois périodes étudiées (1695-1798 ; 1799-1872 et 1873-1920). Toutefois, la productivité la plus élevée de ces composés (P=0.044) ne correspond pas à la taille de la famille morphologique la plus élevée (F=102). Les composés N-de-N sont les plus productifs pendant la première période (P=0.044) ; cependant, la taille catégorielle de leur famille morphologique était la plus basse à cette époque-là.

6. Conclusion et perspectives

La recherche menée sur les noms composés N-de-N a estimé la productivité quantitative de ce type de composés au moyen de deux mesures statistiques différentes : la mesure des hapax (Baayen et Lieber 1991; Baayen 1992; Hay et Baayen 2003) et la mesure de la taille catégorielle morphologique (Baayen et Hay 2002).

Cette étude a montré que les composés N-de-N peuvent être considérées comme productifs dans le corpus analysé : la productivité moyenne de ce type de composés à travers quatre périodes analysées est $P=0.031$. Nous avons aussi remarqué que les composés ayant de grandes familles morphologiques possèdent de la haute degré de la productivité : $F=90/P=11.66$ vs $F=102/P=13.56$; etc.

Un des buts de la recherche visait à vérifier la corrélation potentielle entre les deux mesures de productivité pour les composés N-de-N. Nous avons avancé l'hypothèse qu'il existe une corrélation directe entre la productivité et la taille de la famille morphologique des composés *N-de-N* : la croissance de la taille de la famille morphologique doit augmenter le niveau de leur productivité.

Les résultats présentés dans le Tableau 4 montrent que, globalement, dans le cas des composés N-de-N, il y a une corrélation directe entre la productivité et la taille catégorielle de la famille morphologique. Cette corrélation est évidente pour la 2^{ième}, la 3^{ième} et la 4^{ième} périodes où la croissance de la taille catégorielle de la famille morphologique (T2= 1.85 ; T3=1.95 ; T3=2.01) augmente le niveau de la productivité des composés *N-de-N* (P2=0.017 ; P3=0.029 ; P4=0.032).

Il est intéressant de remarquer que la corrélation ne tient pas pour la première période. L'absence de corrélation entre la productivité et la taille catégorielle de la famille morphologique dans les composés N-de-N peut être expliquée par les conditions linguistiques spécifiques en France à cette époque-là. Le fait que le latin soit moins fréquent et commençait à perdre son importance a permis à la langue française de former de nouveaux mots et d'être largement introduite dans la vie quotidienne à travers plusieurs néologismes.

Pour vérifier une corrélation possible entre la productivité et la taille morphologique des noms composés français, des recherches complémentaires sur d'autres types de composés sont nécessaires aussi bien que des mesures statistiques plus raffinées pour établir les corrélations. Ce travail permettra de

compléter la description du processus de la productivité des noms composés en français qui reste en général peu étudié.

Références

- Baayen, R. H., Lieber, R. 1991. Productivity and English derivation: a corpus-based study, *Linguistics*, 29, pp.801-43.
- Baayen, R.H. 1992. Quantitative aspects of morphological productivity, *Yearbook of Morphology*, Kluwer Academic Publishers, , pp.109-149.
- Baayen, R.H. 1993. On frequency, transparency and productivity, *Yearbook of Morphology*, Kluwer Academic Publishers, pp.181-208.
- Baayen, H., Renouf A. 1996. Chronicling the Times: productive lexical innovations in an English newspaper, *Language* 72: 69-96.
- Bauer, L. 2001. *Morphological productivity*, Cambridge: Cambridge University Press.
- Dal, G. 2003. Productivité morphologique : définitions et notions connexes, dans *Langue française*, v.140, Cedex, Larousse, Paris, pp.3-23.
- De Jong, N. H., Schreuder, R. and Baayen, R. H. 2000. The morphological family size effect and morphology, *Language and Cognitive Processes* 15, 329–365.
- De Jong, N.H., Schreuder, R., and Baayen, R.H. 2003. Morphological resonance in the mental lexicon. In *Trends in Linguistics: Morphological Structure in Language Processing*, eds. R.H. Baayen and R. Schreuder, 65-88. Berlin, Germany: Mouton de Gruyter.
- del Prado, M., Bertram, R., Haikio, T., Schreuder, R., Baayen, H. 2004. Morphological Family Size in a Morphologically Rich Language: The Case of Finnish Compared With Dutch and Hebrew, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 30, No. 6, p.1271–1278.
- Dijkstra, T.; del Prado, M.; Schulpen, B.; Schreuder, R.; Baayen, H.; 2005, A roommate in cream: Morphological family size effects on interlingual homograph recognition, in: *Language and Cognitive Processes*, Volume 20, Issue 1 & 2 , pages 7 – 41.
- Fernández-Domínguez, J., Diaz-Negrillo, A., Pavol Štekauer, P., 2007. How is Low Morphological Productivity Measured?. *Atlantis* 29.1: 29-54.
- Hay, J., Baayen, H., (2002). Affix Productivity and Base Productivity. Paper presented at the Morphological Productivity Seminar, ESSE 6, Aug 30-Sept 3, Strasbourg.
- Krott, A., Schreuder, R., Baayen, R. H.1999. Complex words in complex words, *Linguistics*, 37-5, pp. 905-926.
- Schreuder, R. and Baayen, R. H.1997. How complex simplex words can be, *Journal of Memory and Language* 37, 118–139.

Les corpus

- Frantext :University of Chicago. ARTFL Project.
<http://humanities.uchicago.edu/orgs/ARTFL/>
- Le Dictionnaire Le Littré, Les monuments historiques de la langue française sur CD-Rom, Redon