



HAL
open science

Productivité morphologique : définitions et notions connexes

Georgette Dal

► **To cite this version:**

Georgette Dal. Productivité morphologique : définitions et notions connexes. Langue française, 2003, La productivité en questions et en expérimentations, 140, pp.3-23. hal-01476794

HAL Id: hal-01476794

<https://hal.univ-lille.fr/hal-01476794v1>

Submitted on 16 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open licence - etalab

Productivité morphologique : définitions et notions connexes

In: Langue française. N°140, 2003. pp. 3-23.

Abstract

Georgette Dal : Morphological productivity: definitions and related concepts

This is a state of the art article on the question of productivity. It first discusses the various definitions of productivity that have been proposed, and then examines the measures that have been suggested, with special attention to those proposed by Baayen. This survey leads to two results. The first is that the main recent developments do not concern the concept of productivity itself, but rather new measuring procedures which integrate both qualitative and quantitative aspects. The second is a significant increase in the importance of corpus-based studies in constructional morphology. Complex words are observed in situ in naturally occurring environments. Corpus-based studies allow one to distinguish active and inactive areas in a language, and to avoid confusing central and peripheral types of data.

Citer ce document / Cite this document :

Dal Georgette. Productivité morphologique : définitions et notions connexes. In: Langue française. N°140, 2003. pp. 3-23.

doi : 10.3406/lfr.2003.1063

http://www.persee.fr/web/revues/home/prescript/article/lfr_0023-8368_2003_num_140_1_1063

Georgette DAI.

UMR 8528 « SILEX » (Un. Lille 3 & CNRS) et GDR 2220

dal@univ-lille3.fr

PRODUCTIVITÉ MORPHOLOGIQUE : DÉFINITIONS ET NOTIONS CONNEXES¹

Introduction et présentation du numéro

Le nom *productivité* est l'un de ces termes linguistiques dont l'emploi est miné. Bien qu'il appartienne à ce que j'appellerai le bagage linguistique partagé, et peut-être justement à cause de cela, on ne prend pas toujours le soin de le définir, oubliant que c'est un terme de métalangage, y compris là où on s'attendrait à ce qu'il le soit (par exemple, Arrivé & al. (1986) l'utilisent dans l'article **suffixes**, mais ne lui consacrent pas d'entrée spécifique ; de même, il ne reçoit aucune définition dans la présentation de Dugas & Molinier édés (1992), pourtant intitulé *La productivité lexicale*). Et, quand il est défini, il ne l'est pas toujours de la même façon, à tel point que, selon la définition qu'il reçoit, on peut considérer le phénomène qu'il recouvre comme tout à fait central ou tout à fait marginal, et ce, dans le seul champ de la morphologie constructionnelle².

La notion de productivité est pourtant ancienne : d'après Kastovsky (1986), Jespersen (1942) est le premier à y avoir recouru, pour ce qui est de l'anglais, tandis que Schultink (1992 : 188) va plus loin en soutenant qu'elle est déjà en filigrane chez les grammairiens du sanskrit³. Mais elle est peu explorée, en France du moins. Alors qu'elle vient de se voir consacrer coup sur coup deux ouvrages entiers hors de France et qu'elle fait l'objet de nombreux travaux aux Pays-Bas depuis une bonne dizaine d'années, la question de la productivité inspire peu les chercheurs français, que ce soit d'un point de vue théorique – la seule amorce de réflexion théorique menée en France est à ma connaissance due à D. Corbin (cf. Corbin D. 1976 et 1987) et à P. Corbin (cf. Corbin P. 1982) –, ou d'un point de vue expérimental.

1. Je remercie vivement l'ensemble des auteurs du présent numéro qui, grâce aux discussions et aux échanges que nous avons eus pendant plus de deux ans, ont largement contribué à la rédaction de cet article. Un grand merci également à Pierre Corbin et aux relecteurs de *Langue Française*, avec une mention spéciale pour Anne-Marie Berthonneau, pour leur lecture attentive. Il est inutile de préciser que les erreurs ou incohérences résiduelles sont entièrement de mon fait.

2. Suivant D. Corbin, la morphologie constructionnelle englobe toutes les opérations morphologiques construisant des unités lexicales : dérivation, composition, conversion.

3. Le terme, lui, est dû à F. Diez, dans son ouvrage de 1838.

L'objectif du présent numéro⁴ est précisément de tenter de combler en partie ces lacunes :

– Le présent article propose un état de l'art sur la question : après une recension commentée des définitions que la notion de productivité a reçues (§ 1.), il discutera des principales mesures qui ont été proposées (§ 2.), en s'attardant tout particulièrement sur les propositions de H. Baayen, qui, à ce jour, demeurent les plus usitées.

– Les articles suivants sont moins généralistes. L'article de F. Meunier se situe dans le champ de la psycholinguistique. Elle y fait apparaître que les mots construits nouveaux mettant en œuvre des procédés constructionnels productifs constituent la pierre d'achoppement de la plupart des modèles de reconnaissance de mots, quel que soit le mode d'accès lexical qu'ils prônent⁵. Les deux articles suivants traitent chacun de la productivité d'un procédé constructionnel du français, articulant théorie et expérimentation tout en suivant des démarches opposées : S. Aliquot-Suengas part de la productivité de la forme *-ade* apparaissant dans des innovations lexicales pour statuer sur le nombre de suffixes *-ade* du français ; B. Fradin et N. Hathout tirent au contraire parti d'une étude linguistique de la suffixation en *-et(te)* pour en évaluer la productivité au moyen des outils de mesure proposés par H. Baayen. Les deux derniers articles exploitent chacun à sa manière le développement exponentiel des corpus textuels électroniques. F. Namer décrit dans un premier temps une chaîne de traitement de corpus réutilisable pour la préparation et la création d'une base de données lexicales pour le français, et l'utilise dans un second temps pour reproduire sur le français une expérience relatée dans Krott & al. (1999), selon laquelle il existe en allemand et en néerlandais une corrélation entre la productivité quantitative d'un procédé et la complexité des bases sur lesquelles il s'applique. N. Grabar et P. Zweigenbaum observent la productivité d'une partie des procédés formant des adjectifs dénominaux en français à travers divers domaines et genres : ils commencent par comparer langue générale et langue de spécialité (médecine), et s'intéressent ensuite à la variation de la productivité de ces mêmes procédés à l'intérieur d'une même spécialité médicale, en comparant comptes rendus hospitaliers et sites web.

4. Ce numéro constitue le premier aboutissement d'une réflexion menée depuis septembre 2000 dans le cadre de l'action 5 « Bases de données constructionnelles et productivité des procédés morphologiques en corpus » du GDR 2220 « Description et modélisation en morphologie ». Il réunit les membres de cette action, auxquels s'est jointe S. Aliquot.

5. L'ensemble des articles de ce numéro utilisera les termes *mot(s)* et *mot(s) construit(s)* comme si les notions véhiculées par ces termes étaient claires.

1. Définir la productivité : une ou des productivités ?

1.1. Productivité = productivité constructionnelle ?

Alors que la question de la productivité est absente des préoccupations des syntacticiens, elle est un thème récurrent en morphologie constructionnelle :

- elle constitue un passage obligé dans les ouvrages du domaine, qui tous ou presque lui consacrent au moins un paragraphe quand ce n'est pas leur prologue⁶,
- les travaux de linguistique dont le titre intègre le nom *productivité* ou l'un de ses équivalents se situent tous dans ce champ de la recherche,
- un certain nombre de dictionnaires de terminologie linguistique, surtout français⁷, restreignent le champ d'application de la notion au lexique construit (par ex. Mounin éd. 1974, Dubois & al. 1999),
- quand ils abordent la notion, les ouvrages à visée pédagogique présentant les sous-domaines de la linguistique le font systématiquement dans la section consacrée à la morphologie constructionnelle (par ex. Lehmann & Martin-Berthet 1998 : 110-2, Chiss & al. 2001 : 32).

Il existe deux façons antagonistes d'interpréter les observations qui précèdent : soit on en déduit que la notion n'a de pertinence qu'en morphologie constructionnelle, soit qu'elle est particulièrement visible et, peut-être, particulièrement épineuse dans le lexique construit. C'est cette dernière interprétation que je ferai. Mon propos n'est pas ici d'opposer la morphologie constructionnelle et la syntaxe en relayant l'idée que dénoncent Di Sciullo & Williams (1987 : 4) que seule cette dernière est productive, mais seulement de faire remarquer que la productivité ne peut émerger que sur fond de non-productivité. Or, en matière de repérage des schémas non productifs, les dictionnaires jouent un rôle déterminant, en ceci que c'est parce que nous disposons de dictionnaires que nous sommes capables (ou croyons l'être) de repérer les patrons non productifs. En revanche, la syntaxe ne disposant pas de répertoires comparables aux dictionnaires en ce sens que chaque énoncé produit est supposé nouveau, la productivité n'y dispose pas de faire-valoir. Elle n'en est pas pour autant une notion non pertinente dans le domaine (pour une comparaison entre productivités morphologique et syntaxique, cf. Bauer 1983 : 65-74 et 2001 : 213-222).

1.2. Productivité qualitative ou quantitative ?

On peut répartir les différentes définitions que la notion de productivité constructionnelle (à laquelle renverra désormais le seul nom *productivité*) a reçues en trois types, selon qu'elles l'appréhendent sous un angle qualitatif, sous un angle quantitatif, ou qu'elles conjoignent qualité et quantité.

6. Cf. entre autres et par ordre chronologique Aronoff (1976 : 35-45), Bauer (1983 : 62-99), Scalise (1984 : 157-8), Corbin (1987 : 176-8), Di Sciullo & Williams (1987 : 7-8), Matthews (1991 : 69-80), Lieber (1992 : 1-9), Booij (2002 : 9-12). La question de la productivité fait aussi l'objet d'un article de Spencer & Zwicky eds 1998, et de deux articles de Booij & al. eds 2000.

7. Les dictionnaires de terminologie linguistique anglo-saxons (par ex. Crystal 1992) définissent en revanche d'abord ce que recouvre la notion en syntaxe, l'assimilant à celle de compétence syntaxique, avant de l'étendre, le cas échéant, à la morphologie.

1.2.1. Approche qualitative

Une première série d'auteurs conçoit la productivité comme l'aptitude d'un procédé à former de nouvelles unités lexicales, indépendamment de toute autre considération. Cette première définition, que Di Sciullo & Williams (1987 : 8) et van Marle (1992 : 151) s'accordent à qualifier de basique, apparaît dès 1951 dans les travaux de H. Marchand, qui voulait rompre avec la perspective historico-étymologique de ses prédécesseurs.

Cette première définition, qui voit dans la productivité la propriété de procédés, donc de patrons réglés ou pour le moins réguliers⁸, permet en première approximation de tracer une ligne de démarcation entre la notion de productivité et celle de créativité morphologique, définie comme la création de nouvelles unités lexicales sans recourir à des règles⁹. En cela, la notion intéresse au premier chef les morphologues puisqu'elle rencontre les notions de possible en langue (Koefoed & van Marle 2000 : 303), et de compétence (Dressler & Ladányi 2000 : 104).

Selon cette première approche, la notion coïncide avec ce qu'après Corbin P. (1982), Corbin D. (1987 : 42 et 177) a appelé *disponibilité* (traduit par *availability* dans Plag 1999 et Bauer 2001), et, en toute rigueur, ne devrait avoir de pertinence qu'en dehors des dictionnaires de langue, du moins si l'objectif est d'évaluer la productivité d'un patron au moment où est menée l'étude. Par exemple, selon cette définition, la présence hors dictionnaires de *briochu*, *cuissu*, *culu*, etc. que citent Aurnague & Plénat (1997 : 23) suffit à décréter productif en français contemporain le suffixe *-u* formant des adjectifs marquant la possession inaliénable¹⁰. Les dictionnaires peuvent en revanche jouer un rôle s'il s'agit d'évaluer la productivité d'un patron à une époque révolue, même très proche. Par exemple, une recherche rapide dans la dernière version électronique du *Nouveau Petit Robert* (désormais, *PRE*) permet de supposer que le suffixe *-able* a été apte à former des adjectifs à partir de noms de propriété et de sentiment – qu'il a été *productif* dans la présente acception – entre les XII^e et XVI^e siècles, comme l'indiquent les adjectifs *charitable*, *effroyable*, *équitable*, *pitoyable*, tous datés de cette période. Il ne l'est plus en ce tournant de siècle avec ce type sémantique de bases, que l'on se fonde sur l'intuition ou sur des corpus textuels contemporains, qui ne fournissent aucun nouvel adjectif en *-able* dérivé d'un nom de propriété.

Ainsi définie, la notion se heurte toutefois au problème de l'apparition sporadique de nouveaux dérivés mettant en œuvre des procédés réputés ne pas être productifs. À titre d'exemple, je citerai le cas du suffixe anglais *-th*. Bien

8. La notion de régularité est elle aussi problématique. Pour un aperçu, cf. Poitou (1997), Bauer (2001 : 54-6), Fradin (2003, chap. 7).

9. Le terme *créativité* connaît lui aussi des usages qui divergent selon les auteurs (pour un aperçu, cf. Bauer 2001 : 62-71). Je m'en tiens ici à la définition la plus consensuelle, qu'on trouve entre autres dans Lyons (1977 : 549).

10. Si la possession est banale, attendue pour la catégorie d'entités nommées par le nom recteur, l'adjectif indique en outre que le possesseur se distingue par la taille ou le nombre de l'entité possédée (*bébé fessu* ; *arbre branchu*). Si elle est distinctive, l'adjectif indique la seule possession (*homme barbichu*).

qu'on puisse voir en lui un procédé régulier¹¹ (les mots en *-th* sont majoritairement des noms de propriétés comportant un adjectif, généralement monosyllabique, en position de base), que les noms en *-th* attestés soient en outre pour la plupart au moins sémantiquement transparents¹², ce suffixe passe pour être non productif en anglais depuis plusieurs centaines d'années (Zimmer 1964, Aronoff & Anshen 1998, Plag 1999, Bauer 2001). Malgré cela, on observe l'apparition sporadique de néologismes en *-th*, par ex. *coolth*, cité par Aronoff & Anshen (1998 : 243). Avec cette première définition, on se retrouve ainsi dans la position inconfortable de devoir considérer productifs des procédés jugés par ailleurs non productifs.

1.2.2. Approche qualitatif-quantitative

Une seconde définition, qui fait office de standard chez beaucoup de linguistes, ajoute à la précédente deux propriétés, dont l'une se situe dans l'ordre du quantitatif : dans cette seconde acception, un procédé est productif s'il permet de forger de façon non intentionnelle un nombre en principe infini de formations.

Cette définition, qu'on fait remonter à Schultink (1961)¹³, suscite un certain nombre de commentaires portant respectivement sur la non-intentionnalité et l'infinitude des dérivés formés (on trouve également une discussion de ces deux critères dans Plag 1999 : 13-16).

Non-intentionnalité. La précision selon laquelle les dérivés formés par un procédé productif doivent l'être de façon non intentionnelle est, en première analyse, intéressante. De fait, on peut se demander ce qu'à soi seuls révèlent de la productivité des suffixes *-aille* et *-erie* les noms *sauçaille* sous (1) et *gentleman-cambrioleurerie* sous (2), manifestement construits par leurs auteurs de façon intentionnelle (j'emprunte le premier à San Antonio, le second au CD Rom des archives de *Libération*) :

- (1) [...] quant au canard, il baignait dans une saucaille aqueuse en compagnie de pêches en conserve.

11. Tous les linguistes s'accordent à dire qu'un patron productif est nécessairement régulier, Allen (1978) allant jusqu'à identifier la productivité d'un procédé à la régularité de ses produits. Cependant, si elle est une condition nécessaire, la régularité n'est pas une condition suffisante (cf. Corbin 1976, Plag 1999, Dressler & Ladányi 2000, Kiefer 2000, Bauer 2001).

12. D'un point de vue sémantique, la notion de transparence, surtout utilisée par les psycholinguistes, correspond à celle de (sentiment de) compositionnalité : un mot cesserait d'être transparent – deviendrait opaque –, dès qu'on ne percevrait plus de rapport entre sa forme et son sens, même si, au moment de sa construction, ce rapport existait (Baayen & Lieber 1997 : 283-4 opposent de la sorte *judgement* et *department*). Certains (par ex. Frauenfelder & Schreuder 1992) considèrent que la transparence, qu'elle soit formelle ou sémantique, détermine la productivité. D'autres (par ex. Koefoed & van Marle 2000 : 306) voient en elle une condition nécessaire, mais non suffisante. Pour d'autres enfin (par ex. Baayen 1993), elle est un facteur favorisant, mais non obligatoire.

13. « Onder produktiviteit als morfologisch fenomeen verstaan we dan de voor taalgebruikers bestaande mogelijkheid [...] onopzettelijk een in principe niet telbaar aantal nieuwe formatives te vormen » [Par productivité en tant que phénomène morphologique, nous entendons la possibilité, pour les usagers d'une langue, de former de façon non intentionnelle un nombre en principe infini de nouvelles formations (GD)].

- (2) Maniérisme luxuriant des derniers chefs-d'œuvre baignant dans un post-cinéma dorénavant étouffant, [...] « *gentlemancambrioleurerie* » insignifiante avec Cary Grant et Grace Kelly [...].

Ce critère de non-intentionnalité, qu'on trouve déjà dans Pichon (1942), soulève toutefois plusieurs questions, si, du moins, on l'applique au pied de la lettre en considérant qu'il s'agit là du point de vue de l'émetteur :

– S'il est raisonnable de penser que *sauçaille* et *gentlemancambrioleurerie* ont été formés intentionnellement par leurs auteurs, cette décision ne va toutefois pas toujours de soi, soit que, contrairement à San Antonio, leur auteur n'est pas connu pour sa propension à jouer avec les mots, soit qu'on ne dispose pas d'indices (textuels ou paratextuels comme les guillemets en (2)) permettant de trancher. Comment en effet savoir si les huit occurrences d'*étanchéifier* que comptent les archives de *Libération* ont été créées délibérément, ou si les journalistes qui ont utilisé ce verbe, absent des principaux dictionnaires actuels, l'ont cru attesté (ou ne se sont posé aucune question) ?

– Considérer que les mots forgés intentionnellement ne disent rien de la productivité des procédés utilisés implique par ailleurs que la notion ne soit pas pertinente pour au moins deux pans importants du lexique construit :

(i) on évince d'emblée tous les dérivés porteurs d'un marqueur d'évaluation comme les suffixes *-asse* et *-et(te)* dont le rôle peut précisément être de marquer le point de vue du locuteur. Il s'ensuit que, si l'on applique ce critère au pied de la lettre, jamais on ne pourra statuer sur le caractère productif ou non productif de ces suffixes. Or, on peut y recourir pour former de nouveaux mots, comme le montrent les dérivés non conventionnels relevés dans *Libération* tels que *flicasse* ou *barbouzette* ;

(ii) on exclut de même l'ensemble des dénominations des langues de spécialité : les termes créés dans les technolèctes le sont de façon intentionnelle, puisque leur raison d'être est de satisfaire un besoin dénommatif. Ainsi, alors que les langues de spécialité recourent volontiers au suffixe *-ité*, tout particulièrement avec une base en *-able* (par ex., *accidentabilité* ou *assurabilité* relevés dans les fiches documentaires de *Libération*), l'application stricte du critère en cours d'examen interdit qu'on les prenne en considération si l'objet de la discussion est la productivité de *-ité*¹⁴.

Si l'on applique *stricto sensu* ce critère de non-intentionnalité, on aboutit ainsi au paradoxe que, dès lors qu'un locuteur cherche le mot construit approprié (à la situation énonciative, à un besoin dénommatif), alors, il n'utilise pas productivement le procédé auquel il a recouru.

Le paradoxe se résout en partie si, comme notamment Baayen & Lieber (1991 : 808), par « non-intentionnalité », on entend « indistinguabilité ». Dans ce

14. Selon Pichon (1942), seuls sont « vivants » (nous dirions « productifs ») les procédés participant de la dérivation qu'il appelle « spontanée ». La « dérivation fabricative » (« néologie dénomminative » dans les termes de Guilbert 1975 : 40), qui « consiste en ce qu'un technicien donne, par la pleine activité de sa volonté réfléchie, un nom à une notion scientifiquement individualisée » (1942 : 6) n'est, pour lui, qu'« à la rigueur » vivante.

cas, un procédé serait productif quand il serait apte à former de nouveaux mots non perçus comme tels par le récepteur, voire par l'émetteur. Ainsi, *accidentabilité* ou *assurabilité* témoigneraient tout autant de la productivité de *-ité* que *agglutinabilité* ou *annulabilité* attestés, eux, dans les dictionnaires, puisque rien ne permet *a priori* au locuteur de distinguer entre les deux groupes. En revanche, un dérivé comme *coolth*, qui, au moment de sa construction, a vraisemblablement attiré l'attention des anglophones qui l'ont entendu pour la première fois, n'illustrerait pas la productivité de *-th*. Le problème demeure en revanche entier pour les marqueurs d'évaluation, puisque leur rôle peut précisément être d'attirer l'attention du destinataire. Cette interprétation se heurte en tout état de cause au caractère flou de la notion d'indistinguabilité : comme le remarque Plag (1999 : 14), la conscience morphologique des locuteurs est éminemment variable, et un mot peut passer inaperçu chez l'un et être repéré comme nouveau chez l'autre (même quand il est installé de longue date dans le lexique attesté).

Infinitude des dérivés. Dans la définition en cours d'examen, pour être productif, un patron constructionnel doit en outre être apte à produire un nombre en principe infini de dérivés. À son tour, cette seconde exigence se heurte à au moins une difficulté. Tous les procédés de construction de mots imposent des contraintes aux bases qu'ils sélectionnent : contraintes catégorielles, mais aussi sémantiques, phonologiques ou structurelles. Dès lors, le nombre de bases possibles pour un procédé donné peut se trouver limité, si bien que le nombre de dérivés l'est aussi par ricochet : en tout état de cause, ce dernier n'est pas toujours infini, fût-ce « en principe ». C'est particulièrement vrai pour les procédés qui, comme les suffixes *-u* ou *-ifi(er)*, sélectionnent de façon privilégiée des bases non construites. On peut en effet faire l'hypothèse que le nombre de bases non construites qui satisfont en outre aux contraintes sémantiques, catégorielles et phonologiques que de tels procédés imposent aux bases qu'ils sélectionnent – par exemple, tendanciellement pour *-u* : être un nom monosyllabique (Aurnague & Plénat 1997 : 24) de partie/constituant inaliénable – n'est pas infini. Même si de tels procédés forment, actuellement encore, quelques mots, rien ne permet de dire qu'il existe encore une infinité de bases possibles, donc que le nombre de néologismes en *-u* et *-ifi(er)* est lui aussi infini.

1.2.3. Approches quantitatives

À côté des deux définitions qualitatives qu'on vient de voir, il existe dans la littérature plusieurs façons de considérer la notion de productivité sous un angle proprement quantitatif, selon que la quantité concerne celle des dérivés attestés et/ou possibles, ou celle des bases.

L'emploi de l'épithète *productif* pour qualifier un procédé constructionnel ayant produit ou produisant beaucoup de dérivés est certainement le moins linguistique de tous. On trouve pourtant une définition de ce type dans Mounin éd. (1974 : s.v. **productivité**), pour qui « un procédé de création lexicale, par exemple la dérivation affixale, est dit productif lorsque les formes nouvelles auxquelles il donne ou a donné naissance sont nombreuses ». C'est également ainsi que l'emploi de façon non contrôlée un certain nombre d'études linguistiques. Or, même s'il peut être utile de connaître ce que Corbin (1987 : 42)

nomme la « rentabilité » d'un procédé (*profitability* dans les termes de Plag 1999 et Bauer 2001), à soi seule, cette définition ne définit pas la productivité des procédés mais leur production effective puisque, dans ce cas, la productivité n'est plus une aptitude. Utilisée de façon brute, elle confond en outre productions présente et passée, et ne dit en définitive rien linguistiquement du procédé concerné, si ce n'est qu'il figure dans peu ou dans beaucoup de mots. Enfin, comme Corbin (1987 : 42), on reprochera à cette approche de s'appuyer sur des inventaires nécessairement fluctuants et soumis aux aleas de l'attestation.

Cette définition, qui prend en compte le nombre de dérivés effectivement produits, a toutefois donné lieu à une variante, qu'on trouve notamment chez Zwanenburg (1983 : 28). Selon lui :

[o]n peut [...] définir la productivité d'une règle dérivationnelle comme le quotient du nombre de mots que cette règle peut former par le nombre de mots qu'on rencontre réellement dans la langue.

Cette variante présente l'avantage de prendre en compte à la fois le possible (correspondant au dividende du quotient) et l'effectif (correspondant au diviseur) : la productivité redevient dans ce cas une notion linguistique, puisqu'elle fait de nouveau appel à l'aptitude de la règle à former de nouveaux mots. Elle pose malheureusement plusieurs problèmes, dont l'un à mon sens rédhibitoire. En effet, pour pouvoir donner une valeur au dividende du quotient – en admettant que cela soit possible –, il faut au préalable avoir déterminé la productivité de la règle par un autre moyen. Autrement dit, en dépit du verbe *définir* qu'emploie W. Zwanenburg, il ne s'agit pas là d'une définition de la productivité mais d'une estimation chiffrée.

À côté des définitions quantitatives axées sur le nombre de dérivés, on trouve des définitions faisant état de la quantité des bases auxquelles peut s'appliquer un procédé donné (cf. par ex. Trask 1993 : s.v. **productive**, où le suffixe *-ness* anglais est dit productif parce qu'il peut s'appliquer à (presque) n'importe quel adjectif¹⁵). De telles définitions posent un problème qu'on a déjà souligné : elles supposent en effet que les procédés constructionnels s'appliquent dès que sont satisfaites les conditions catégorielles propices à leur application. Or, même si on peut déplorer quelques dissonances ici ou là (cf. Dal 2002), il est désormais banal d'admettre que les procédés de construction de mots font peser autre chose que des contraintes catégorielles sur les bases qu'ils sélectionnent.

De fait, en 1976 déjà, M. Aronoff faisait observer que, sauf dans de rares cas, la productivité d'une règle de construction de mots était toujours restreinte à une sous-classe d'une classe syntaxique de bases. Cette remarque a inspiré une approche de la productivité définie en substance comme l'aptitude d'un procédé à contraindre faiblement les bases sur lesquelles il s'applique. On trouve cette façon de concevoir la productivité dans Booij (1977 : 5) et, plus récemment, (2002 : 101) : dans ces deux travaux, d'un point de vue quantitatif, la productivité d'un procédé est posée comme inversement proportionnelle au nombre de restrictions portant sur les bases. À son tour, cette définition pose

15. Cf. aussi Koefoed & van Marle (2000 : 302).

problème : en effet, il existe des procédés peu contraignants envers les bases et pourtant inaptes à former de nouveaux mots (Aronoff 1976 et Scalise 1984 : 157 citent le cas du suffixe *-ous* anglais) ; inversement, un suffixe comme *-u* a déjà montré qu'un procédé peut contraindre fortement les bases qu'il sélectionne tout en étant apte à former de nouvelles unités lexicales.

1.3. Productivité absolue ou relative ?

Avant de clore cette première partie, on se demandera rapidement s'il faut voir dans la productivité une notion binaire ou une notion scalaire. De nouveau, on renverra dos à dos les approches qualitatives et quantitatives de la notion¹⁶ :
– que ce soit dans l'approche qualitative pure (§ 1.2.1.) ou dans celle qui conjoint qualité et quantité (§ 1.2.2.), la notion est binaire : un procédé constructionnel ne peut être qu'apte ou inapte à former de nouveaux mots. C'est ainsi que, pour Zwanenburg (1983 : 29-30), qui se réclame de Halle (1973) et de Corbin (1976), il existe une « opposition discrète entre procédés productifs et improductifs » (on retrouve la même opposition dans Booij (1977) et Corbin (1987 : 177-8,) à ceci près que, pour elle, l'opposition trace une ligne entre processus disponibles et non disponibles) ;
– en revanche, les diverses approches quantitatives qui ont cours ont en commun de voir dans la productivité une notion scalaire susceptible, selon les auteurs, soit de connaître seulement trois degrés (dans cette tripartition, un procédé est alors productif, non productif ou semi-productif, cette dernière notion variant elle-même selon les auteurs : pour une synthèse, cf. Bauer 2001 : 16-8), soit d'être affectée d'un indice chiffré : la notion est alors conçue comme un continuum (Aronoff & Anshen 1998 : 243), pouvant aller du non productif à l'entièrement productif, en passant par toutes les valeurs intermédiaires.

2. Mesurer la productivité

On s'attendrait à ce que seuls les adeptes d'une approche proprement quantitative de la notion cherchent à mesurer la productivité des procédés constructionnels. En effet, dans l'approche strictement qualitative vue au § 1.2.1., il ne devrait pas y avoir de sens à développer des calculs mesurant la productivité, puisque la découverte d'une seule unité lexicale construite ne figurant pas dans les dictionnaires suffit pour décréter productif en synchronie le procédé qui l'a formée. De la même façon, même si ce n'est pas pour les mêmes raisons, dans l'approche qualitatif-quantitative exposée au § 1.2.2., mesurer la productivité devrait être non pertinent, puisque, utilisée intuitivement, la notion d'infini s'oppose à celle de dénombrabilité.

Les choses ne sont toutefois pas aussi tranchées : on trouve en effet des linguistes qui se situent dans une perspective qualitative de la notion et qui ont

16. Dressler & Ladányi (2000) font exception : tout en adoptant explicitement la définition qualitatif-quantitative de H. Schultink, ils distinguent cinq degrés de productivité fondés sur des critères qualitatifs. Cette apparente incongruité s'explique par le cadre de la Morphologie Naturelle dans lequel ils s'inscrivent, où les concepts sont conçus comme graduels.

malgré cela développé des méthodes de calcul statistiques. Le paradoxe n'est finalement qu'apparent, car, comme le souligne Plag (1999 : 22), si quelque chose peut arriver, il doit être possible de mesurer dans quelle probabilité.

2.1. L'indice de productivité d'Aronoff 1976

Aronoff 1976 est le premier parmi les générativistes à avoir proposé d'affecter chaque règle de construction de mots d'un indice de productivité, permettant de comparer la productivité des règles entre elles. Cet indice, qu'on obtient en divisant le nombre de mots effectivement produits par la règle concernée par le nombre de mots qu'elle est apte à former (ce que Baayen & Lieber 1991 : 803 notent par la formule $I = V/S'$, où V correspond au nombre de dérivés attestés porteurs du procédé étudié, et S' au nombre de dérivés que ce procédé peut former), a déjà donné lieu à de nombreuses critiques portant en substance sur la difficulté à déterminer ce que sont l'« actuel » et le « possible » en matière de mots construits – et, par ricochet, sur la difficulté à donner une valeur chiffrée aux produits actuels et possibles d'une règle de formation de mots donnée –, ainsi que sur les résultats contre-intuitifs obtenus pour les cas extrêmes.

2.1.1. L'actuel et le possible en matière de mots construits¹⁷

Si l'on admet qu'un procédé constructionnel est productif quand il permet de former, de façon réglée, de nouvelles unités lexicales sans qu'elles soient particulièrement remarquables, alors il semble raisonnable qu'un calcul de la productivité fasse intervenir à la fois le possible – pour des raisons évidentes – et l'actuel : en effet, pour déterminer si un mot est nouveau, il faut au préalable savoir ce qu'est un mot non nouveau (d'après Bauer 2001 : 34, van Santen 1992 souligne pareillement le caractère crucial de la notion de « mot actuel » pour la notion de productivité). Il y aurait ainsi au moins un domaine où la notion de mot actuel aurait quelque intérêt car, pour certains linguistes (par ex., Kiparsky 1982 : 26), la notion est inintéressante aussi bien d'un point de vue théorique que d'un point de vue empirique (selon eux, aucune règle de formation de mots n'est tributaire du fait qu'un mot existe ou non).

Cependant, malgré son apparente immédiateté, la notion d'« actuel » en matière de mots construits n'a rien de trivial :

– un mot accède-t-il à l'existence quand il entre dans le dictionnaire ? Si oui, le(s)quel(s) (dictionnaires de langue générale ou de spécialité, dictionnaires diachroniques et/ou synchroniques, et, si oui, quelle synchronie ?...)?

– un mot existe-t-il dès lors qu'il appartient au lexique d'un individu donné ? À celui d'un locuteur idéal qui subsumerait celui de tous les individus particuliers ? À celui d'une communauté linguistique circonscrite ? Si oui, laquelle (lesquelles), et toutes les communautés linguistiques sont-elles à traiter sur un pied d'égalité ?

17. La plupart des ouvrages de morphologie discutent de la dichotomie mots possibles/mots actuels (aussi dits *existants*) : c'est notamment le cas de Corbin (1987), Plag (1999 : 6-11), Bauer (2001 : 34-43). Cf. aussi Corbin & al. éd. (1997), intitulé « Mots possibles et mots existants ».

– un mot construit est-il « actuel » dès qu’il apparaît dans un corpus non dictionnaire et, si oui, tous les corpus se valent-ils ?

Pour Aronoff (1976 : 19), il semble que soit actuelle la liste de mots dont un locuteur dispose à un moment donné, en d’autres termes de ce qu’il est convenu d’appeler le « lexique mental » d’un individu X à un temps T. Or, et comme le fait observer Rainer (1987 : 195), cette définition n’est pas consistante avec un calcul de la productivité définie comme le rapport entre le nombre de mots actuels et le nombre de mots possibles, en ceci que l’indice associé à un procédé donné est susceptible de variations interindividuelles mais aussi de variations intraindividuelles à des temps différents : si l’indice peut varier dans l’espace et dans le temps – y compris à des moments très proches –, alors, il est de peu d’intérêt pour la théorie linguistique¹⁸.

La notion de « mot possible » pose, elle aussi, un certain nombre de problèmes. En substance, on peut se demander ce qu’est un mot possible, si l’ensemble des mots possibles inclut ou non celui des mots existants, et, plus crucialement, comment donner une valeur chiffrée au possible.

S’agissant de la première question, on admet en général qu’un mot construit est possible quand il ne contrevient pas aux règles. La notion de « mot possible » est alors à distinguer de celle de « mot probable » : en effet, comme le montrent Corbin (1997) et, après elle, Dressler & Ladányi (2000 : 106), certains mots peuvent être bien formés, et pourtant avoir une probabilité d’actualisation faible, faute de trouver un contexte d’actualisation adéquat ou un référent plausible. Peuvent également faire obstacle à l’actualisation de mots linguistiquement possibles des problèmes de performance (comme la longueur des mots) ou d’acceptation sociale.

En ce qui concerne l’inclusion ou non des mots construits existants dans les mots construits possibles, on relève des positions divergentes dans la littérature linguistique : pour les uns, les mots construits existants sont un sous-ensemble des mots possibles (c’est par exemple la position de Dal 1997a) ; pour d’autres, l’ensemble des mots construits possibles est le complémentaire, dans l’ensemble des mots construits, des mots construits existants, soit parce que, simplement, pour eux, une possibilité n’en est plus une quand elle est actualisée, soit que, dans la veine de Bloomfield (1933), ils considèrent que l’entrée dans le lexique se charge automatiquement d’idiosyncrasie (cf. Fradin 1997 : 101, pour qui les mots existants, construits ou non, sont ceux qui appartiennent au lexique d’une langue compris minimalement comme « le répertoire des unités présentant une ou des propriétés qui ne peuvent être prédites par les règles de cette langue ») ; pour d’autres encore (par ex. Corbin 1997 : 80), le lexique construit actuel compte parmi ses rangs des mots bien formés (donc possibles), et des mots au moins en partie mal formés.

18. On pourrait lui trouver un intérêt si l’on était capable d’effectuer des calculs globaux comparant les indices I d’un individu X à des temps différents, les indices d’individus différents à un temps T, et de croiser les résultats.

On voit que, selon la définition qu'on retient, le calcul de l'indice I d'un procédé donné varie : dans le premier cas, le dividende correspond au nombre de mots existants construits avec ce procédé, et le diviseur ajoute à ce nombre le nombre quantifiant le lexique possible non existant ; dans le second cas, le diviseur ne comprend que le complémentaire de l'existant dans le bien construit ; dans le troisième cas, le dividende estime la quantité existante mal construite présentant le procédé considéré, le diviseur quantifiant les mots bien construits, y compris les mots éventuellement attestés dans les dictionnaires.

Enfin, la quantification du possible, quelle que soit la définition qu'on retienne, suppose quant à elle d'être entièrement au clair avec le procédé dont on calcule l'indice. Elle demande en effet d'abord qu'on détermine l'ensemble des contraintes qui pèsent sur l'application du procédé, qu'on dénombre ensuite le nombre de bases les satisfaisant en réitérant la recherche si la base peut elle-même être construite, et ce autant de fois que la base admet de type de constructions. Appliquée par exemple au suffixe *-ité* français, la démarche revient :

(i) à déterminer précisément à quel type sémantique de bases il peut s'appliquer. Dire qu'elles doivent exprimer des propriétés objectives (Corbin à paraître) ou qu'à leur référent soient associables des propriétés objectives (Dal 1997b) est une première approximation. Mais elle ne suffit pas, ne serait-ce que parce qu'on ne dispose pas de typologie des bases sur laquelle se fonder ;

(ii) à calculer le nombre de mots possibles en *-able, -aire, -el, -eux, -if, -ique* qu'affectionne le suffixe *-ité* si l'on se fonde sur le lexique attesté dans le TLF (cf. tableau 1), donc, à reprendre pour chacun de ces suffixes la procédure à (i), étant entendu que certains d'entre eux sont eux-mêmes applicables à des verbes construits ;

(iii) à déterminer si tous les mots possibles construits par l'un de ces procédés constituent une base possible pour la suffixation en *-ité*.

Base	Nombre	Exemple
Xable	239	<i>acceptabilité</i>
Xel	237	<i>accidentalité</i>
Xif	138	<i>absorptivité</i>
Xique	130	<i>algébricité</i>
Xeux	84	<i>callosité</i>
Xaire	69	<i>alimentarité</i>

Tableau 1. Dérivés en *Xité* selon le type morphologique apparent de la base attestés dans le TLF

2.1.2. Fausses prédictions

Outre la difficulté qu'on vient d'évoquer, l'indice de productivité d'Aronoff conduit à de fausses prédictions pour les cas extrêmes. À la suite de Plag (1999 : 23), on remarquera en effet que les affixes aptes à former un grand nombre de nouveaux dérivés se voient affectés d'un indice tendant vers zéro (puisque le diviseur tend alors vers l'infini), tandis qu'au contraire, l'indice des affixes peu aptes à former de nouveaux mots peut être très supérieur à 1 (puisque le nombre de mots actuels excède celui des possibles), celui des affixes non disponibles étant, lui, incalculable (puisque le diviseur prend alors la valeur zéro)¹⁹.

Les seuls résultats exploitables concernent donc des cas moyens, ce qui minore l'intérêt de la mesure.

2.2. Les mesures d'H. Baayen

Ce sont certainement les défauts mêmes de l'indice de productivité d'Aronoff qui sont à la source des mesures proposées par H. Baayen, puisque, comme on le verra, elles évitent les écueils qu'on vient de souligner.

H. Baayen, dans ses travaux seul ou en collaboration, fait sienne la définition de H. Schultink selon laquelle un patron est productif quand il est apte à former un nombre en théorie infini de dérivés construits de façon non remarquable. La raison pour laquelle il cherche à calculer la productivité des patrons constructionnels n'est pas tant l'intérêt du calcul en soi que le fait que, pour lui, seuls les procédés productifs intéressent la théorie morphologique (Baayen & Lieber 1991 : 801-2)²⁰. En même temps, l'investigation menée s'inscrit dans le champ de la psycholinguistique (Frauenfelder & Schreuder 1992 : 165 le remercie d'ailleurs d'avoir remis la notion de productivité à l'agenda des psycholinguistes).

H. Baayen a ainsi développé deux mesures de calcul de la productivité : la « productivité au sens strict » et la « productivité globale ». Comme la plus usitée est la productivité au sens strict, et comme la productivité globale pose des problèmes méthodologiques analogues à ceux que pose la productivité au sens strict à ceci près qu'ils sont multipliés par le nombre de procédés mis en jeu par le diviseur, c'est d'elle et d'elle seule qu'il s'agira dans ce qui suit²¹.

La productivité « in the strict sense » (Baayen & Lieber 1991 : 817), notée P , est décrite comme une formalisation mathématique de la notion linguistique de productivité morphologique (*ibid.* : 811) exprimant la propension d'un procédé à former de nouveaux mots (Baayen 1993 : 181). Il s'agit d'une mesure relative,

19. À moins qu'il ne soit égal à 1, si l'on considère que les mots actuels sont des mots possibles (Plag 1999 : 23).

20. Cf. aussi Lieber (1992 : 1). Corbin (1976) défend le point de vue strictement opposé : pour elle, les procédés non productifs intéressent aussi le morphologue.

21. La « productivité globale » (« global productivity »), notée P^* , exprime la probabilité, pour un mot entièrement nouveau, qu'il contienne un procédé donné (Hay & Baayen 2002). Dans ses premiers travaux, H. Baayen la représente par un plan orthonormé prenant P en abscisse, et le nombre de dérivés dans lequel apparaît le procédé étudié en ordonnée (noté V , pour *vocabulary*). P^* est ainsi un moyen d'intégrer la rentabilité du procédé considéré 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). Toutefois, sauf dans le cas idéal où P_A et V_A sont supérieurs à, respectivement, P_B et V_B (A est alors clairement plus productif que B), elle pose un problème de lecture, soit que P_A soit supérieur à P_B et V_A inférieur à V_B , ou qu'inversement, P_A soit inférieur à P_B et V_A supérieur à V_B . En réponse à van Marle (1992) qui lui reprochait pareillement ces cas indécidables, Baayen (1993), après avoir justifié l'adéquation empirique de P^* , se livre à des calculs mathématiques sur lesquels je passe, pour finir par proposer la formule simplifiée $P^* = n_{1,E,A} / h_1$, où $n_{1,E,A}$ note le nombre d'hapax porteurs du procédé E pour lequel on calcule P^* , et h_1 le nombre total d'hapax (construits ou non) du corpus (pour une autre notation, cf. Baayen 1994b).

destinée à comparer la productivité de procédés à l'intérieur d'un même corpus (Baayen & Renouf 1996), ou entre corpus différents (Baayen 1994a).

P correspond au quotient du nombre n_1 (parfois noté $V(1,n)$) d'hapax legomena (types ne possédant qu'une occurrence dans le corpus considéré²²) construits au moyen du procédé étudié par le nombre total N d'occurrences des types comportant ce procédé dans le corpus :

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

H. Baayen justifie à plusieurs reprises la raison pour laquelle P se fonde sur le nombre d'occurrences des types et non, plus simplement, sur le nombre de types. L'explication part du constat que les types à haute fréquence prédomineraient pour les procédés inaptes à former de nouveaux mots, tandis que les procédés aptes à former de nouveaux mots se caractériseraient par des types à basse ou moyenne fréquences (cf. notamment Baayen 1993 : 181)²³. Selon ce principe, les types mettant en jeu un procédé non disponible font beaucoup augmenter la valeur du diviseur, donc sensiblement baisser la valeur de P , tandis que les types mettant en jeu un procédé qualitativement productif sont de peu d'effet sur la valeur du diviseur, donc sur celle de P .

Même si elle a été critiquée²⁴, P continue à ce jour de faire office de standard quand il s'agit de comparer de manière chiffrée la productivité de divers procédés (par exemple, après avoir montré l'un et l'autre ses limites, Plag 1999 et Bauer 2001 l'utilisent à leurs propres fins). Elle demande toutefois qu'on clarifie le rapport entre hapax et productivité, et se heurte, dans les faits, à plusieurs problèmes méthodologiques.

22. Un type est une expression qui instancie un patron constructionnel, un même type étant susceptible de connaître plusieurs occurrences dans un corpus donné (par exemple, le type *productivité* compte 96 occurrences dans le présent article).

23. La fréquence est l'une des pistes qu'explorent les psycholinguistes pour expliquer la productivité. En 1993, H. Baayen, qui préconise un modèle psycholinguistique d'accès lexical dans lequel l'accès peut suivre deux chemins, celui de l'analyse ou celui de l'accès direct, considère qu'en situation de reconnaissance, la fréquence de l'analyse, par conséquent celle de l'activation du procédé considéré, influe sur sa productivité, cette fréquence d'analyse étant elle-même corrélée à la fréquence absolue du dérivé (on accéderait directement aux dérivés à haute fréquence, par analyse aux dérivés à moyenne ou basse fréquences). Depuis, Hay (2001) et Hay & Baayen (2002) ont montré qu'indépendamment de la fréquence absolue interviennent les fréquences relatives du dérivé et de sa base : les dérivés plus fréquents que leur base (par ex. en anglais *inaudible* vs *audible*) seraient plus sujets à l'accès direct, les dérivés moins fréquents que leur base (par ex. *inadequate* vs *adequate*) plus sujets à l'analyse. Il s'ensuit qu'on accéderait directement aux dérivés à basse fréquence si leur base est moins fréquente encore, et qu'on décomposerait les formes à haute fréquence si leur base a une fréquence encore plus forte.

24. L'une des critiques qui lui ont été faites est d'être trop indexée sur la performance, et de ne rien dire en définitive de la compétence (cf. van Marle 1992 et Dressler & Ladányi 2000 : 104). Dans plusieurs de leurs travaux récents (par ex. Gaeta & Ricca 2003), L. Gaeta et D. Ricca formulent une autre critique portant sur le fait que N désavantage les affixes très fréquents. Pour pallier cet inconvénient, ils proposent de quantifier par N le nombre total des occurrences du corpus, et de comparer les productivités de procédés à valeurs de N constantes.

2.2.1. Rapport entre hapax et productivité

Étant donné la définition linguistique de la notion de productivité dont *P* entend être une mesure mathématique, l'attente est double : on s'attend d'abord à ce qu'en fait d'hapax, n_1 quantifie des néologismes, *i.e.* des types constituant des innovations lexicales (ou des types considérés comme tels²⁵), ensuite, si l'équivalence hapax/néologisme est prouvée, que les hapax relevés témoignent d'une aptitude.

S'agissant de la seconde attente, van Marle (1992 : 152) considère qu'elle n'est pas satisfaite : pour lui, *P* mesure ce qui est effectivement produit dans le corpus de référence, et non une aptitude. Baayen (1993 : 189) lui répond en substance que le nombre d'hapax sert de révélateur à la probabilité de rencontrer de nouveaux types qui n'ont pas été enregistrés dans l'échantillon précédemment exploré. En tant que mesure probabiliste, *P* serait ainsi projetable sur un corpus de taille étendue à venir.

Si l'on convient que tout calcul de probabilité se fonde sur l'existant, il faut donc s'assurer de l'équivalence hapax/néologisme. La question peut être examinée sous deux angles : est-ce que tous les hapax d'un texte sont des néologismes ? Est-ce que tous les néologismes d'un texte sont des hapax ?

En ce qui concerne l'implication hapax/néologisme, elle ne va pas de soi. Se pose entre autres le problème des mots peu utilisés ou inattendus étant donné le genre textuel du corpus sur lequel se fait le calcul. Par exemple, comme le montre le tableau 2, *fourgonnette* ou *pendulette*, tous deux attestés dans le *PRE* qui les date respectivement de 1949 et 1893, sont hapaxiques dans *Le Monde diplomatique* [LMD] entre janvier 1984 et septembre 1998, mais possèdent plusieurs occurrences dans *Le Monde* [LM] entre janvier 1997 et décembre 1998, et dans *Libération* [libé] entre janvier 1995 et décembre 1999. Le calcul de *P* pour *-ette* dans LMD prendrait donc au titre de n_1 des dérivés bien installés dans le lexique français (et largement représentés dans d'autres corpus textuels) :

	<i>PRE</i>	LMD	LM	libé
clarinette(s)	1753	1	159	147
fourgonnette(s)	1949	1	39	83
pendulette(s)	1893	1	14	17
voiturette(s)	1896	1	21	22

Tableau 2. Comparaison du nombre d'occurrences de quelques types en *-ette* dans trois corpus différents

Il peut même arriver qu'un dérivé hapaxique dans un corpus donné soit justement un mot en passe de disparaître du vocabulaire, autrement dit un archaïsme : c'est par exemple le cas dans *Libération* d'*affriander* ou de *dormitif*, marqués comme « vieux » par le *PRE*.

25. La notion de « néologisme » est, elle aussi, mal cernée. Un mot peut en effet appartenir au lexique conventionnel, sans que le locuteur qui le produit le sache.

Quant à l'implication réciproque néologisme/hapax, elle peut elle aussi être mise en doute. Il arrive que, dans un corpus donné, on trouve deux occurrences ou plus d'un mot absent des dictionnaires – donc d'un néologisme –, par ex., le verbe *contretamponner* présent à deux reprises dans *Libération*. Il arrive même que les diverses occurrences du même mot soient le fait du même auteur. Le cas s'observe de nouveau dans les corpus journalistiques, où certains journalistes ont une rubrique récurrente, dans laquelle ils usent de façon tout aussi récurrente d'un mot qui leur est idiolectal. Ainsi *antiauto(s)*, dont les deux occurrences dans *Libération* ont été produites par la même journaliste à un an d'intervalle (on en notera l'orthographe fluctuante) :

- (3a) Strasbourg se pose en modèle de la lutte antiauto. (N. Gautier, 22/9/1998)
- (3b) La capitale alsacienne a effectivement été l'une des pionnières de la lutte antiautos. (N. Gautier, 20/9/1999)

ou *boyesse*, dont les deux seules occurrences dans *Libération* figurent, elles, dans le même article.

H. Baayen et ses coauteurs ont anticipé la critique (cf. notamment Baayen & Lieber 1991 : 801, Baayen 1993 : 188), et ont mené des expérimentations visant à prouver la corrélation entre hapax et néologisme (c'est l'objet de Baayen & Renouf 1996, qui montrent que la proportion des mots construits absents du *Third New International Dictionary of the English Language* qui figurent dans le *Times* sur une période de 3 ans est plus forte parmi les hapax que parmi les dérivés non hapaxiques). Cependant, l'équivalence hapax/néologisme requiert que le calcul de *P* soit effectué sur un très grand corpus (d'une taille qu'il faudrait du reste spécifier), et que le corpus en question soit échantillonné, du moins si l'objectif est d'estimer la productivité d'un procédé en langue générale. Or, on sait que l'échantillonnage des corpus constitue à soi seul une grosse difficulté.

2.2.2. Problèmes méthodologiques

Le calcul de n_1 et de N pose par ailleurs une série de problèmes méthodologiques. Le premier que je signalerai ne concerne que N , les autres concernent n_1 et N , même s'ils sont peut-être plus visibles pour N (pour d'autres problèmes, cf. Gaeta & Ricca 2003). Les problèmes rencontrés ne sont pas techniques (il existe désormais des systèmes permettant de repérer automatiquement les occurrences d'un texte et de les lemmatiser), mais proprement linguistiques.

Les mots hérités. Pour décider si une occurrence O apparemment porteuse d'un procédé C illustre bien C , il faut déterminer ce qu'on retient au titre de C . En effet, dans une langue comme le français, une bonne partie du lexique apparemment construit est en fait héritée du latin (ou du grec)²⁶. Une première décision concerne donc la prise en compte au titre de N de ces mots hérités. Sauf à la prendre de façon arbitraire, elle nécessite qu'on détermine au préalable si ces héritages pèsent qualitativement sur la productivité des procédés en français actuel, et si oui, si tous pèsent au même titre, selon qu'ils sont analysables en français ou non.

26. Selon Huot (1997 : 141), 80 % des mots du français viennent du latin. Elle ne chiffre toutefois pas la part des mots dérivés parmi ce pourcentage.

À titre d'exemple, considérons le tableau 3, qui donne la répartition des 100 premiers noms féminins de forme *Xation* qu'atteste le *PRE*.

Héritage d'un <i>Xatio</i> latin	Construction en français	Autres	Total
59 (ex. <i>abdication</i>)	36 (ex. <i>acclimatation</i>)	5	100

Tableau 3. Répartition des 100 premiers *Xation* attestés dans le *PRE*

Imaginons qu'on veuille calculer P pour *-ation*, et que, dans le corpus sur lequel s'effectue le calcul, on rencontre l'ensemble des 59 mots en *Xation* de la première colonne : faut-il ou non les comptabiliser ? On voit la répercussion de la décision sur la valeur de N , d'autant que certains de ces types ont en outre une fréquence haute. Le tableau 4 fait ainsi apparaître que, pour me limiter aux 10 premiers de ces 59 *Xation* hérités, P , calculée sur *Libération*, verra son diviseur augmenté ou diminué de 2185 selon le choix opéré :

<i>abdication(s)</i>	60	<i>abomination(s)</i>	78
<i>aberration(s)</i>	284	<i>abréviation(s)</i>	75
<i>abjuration(s)</i>	8	<i>abrogation(s)</i>	439
<i>ablation(s)</i>	107	<i>accélération(s)</i>	925
<i>abnégation(s)</i>	144	<i>accentuation(s)</i>	65
TOTAL			2 185

Tableau 4. Fréquence dans *Libération* des 10 premiers *Xation* hérités du latin selon le *PRE*.

Les homomorphes. Une autre décision non triviale ayant une influence sur le calcul de P concerne les procédés homomorphes. En effet, en français comme dans d'autres langues, il existe des procédés de construction de mots qui ne sont pas apparentés, bien qu'ils aient la même forme. C'est par exemple le cas de la forme *-esse*, qui subsume deux suffixes différents, l'un, désormais *-esse₁*, marquant la sexuisemblance (il marque plus précisément l'appartenance du référent du dérivé au sexe féminin : par ex., *ânesse*, *comtesse*), l'autre, désormais *-esse₂*, apparaissant dans des noms de propriété (*allégresse*, *richesse*).

Dans des cas comme celui-ci, le calcul de P demande donc qu'on opère d'abord un tri selon le suffixe *-esse* qu'illustre l'occurrence relevée, qu'elle le soit au titre de n_1 ou de N : ainsi, les hapax *capitainesse*, *clownesse*, *scoutesse* relevés dans *Libération* seront portés au crédit du calcul de n_1 pour *-esse₁*, tandis que *prestesse* et *simplesse*, aussi issus de *Libération*, serviront au calcul de n_1 pour *-esse₂*.

S'agissant de N toutefois, les données sont plus complexes. En effet, si l'on croise ce problème d'homomorphie avec celui des mots hérités, la situation devient proprement inextricable. Comme le montre le tableau 5, le *PRE* atteste 103 noms féminins de forme *Xesse* : il donne 30 d'entre eux comme construits en français par *esse₁*, 34 comme construits en français par *-esse₂*, 24 comme des emprunts, soit à des mots italiens de forme *Xezza* (*prestesse* < it. *prestezza*, *sveltesse* < *sveltezza*), soit à des mots latins de formes *Xissa* (*diaconesse* < *diaconissa* ; *promesse* < *promissa*) ou *Xicea* (*pesse* < *picea*), soit à des mots de langues variées (espagnol, provençal, anglais, flamand, francique). Sur les 15 mots restants, 8 sont construits

en français, mais la suffixation en *-esse* ne représente pas la dernière étape constructionnelle, les 7 mots résiduels posant des problèmes divers :

Construits en français			Emprunts			Autres
-esse ₁	-esse ₂	autre constr.	lat.	ital.	langues div.	
30	34	8	11	8	5	7

Tableau 5. Noms féminins de forme *Xesse* attestés dans le *PRE*

Les cohortes. Comme Plag (1999 : 28-9) et Gaeta & Ricca (2003), le dernier problème méthodologique que je mentionnerai est celui des cohortes, *i.e.* des affixes pris dans des cycles successifs de dérivation. Le problème est en fait double :

(i) Déterminer si un dérivé de forme *préfxsuf* compte au titre du calcul de *P* pour *préf.* ou pour *suf.* requiert le plus souvent une validation manuelle, ou au moins une analyse préalable déterminant l'ordre d'application des opérations (la remarque vaut aussi bien pour n_1 que pour N)²⁷. J'illustrerai ce cas avec les noms de forme *inXité*. Si la plupart sont analysables comme résultant de l'application de *-ité* à un adjectif en *in-* (ex. *insalubrité*), on peut parfois hésiter, d'autant plus que le lexique attesté comprend quelques noms en *in-* nécessairement dérivés de noms de propriété comme *inharmonie* ou *inconfort*. C'est par exemple le cas d'*inadversité* en (4), relevé comme hapax dans *Libération*, qui exprime davantage l'absence de la propriété attendue qu'exprime *adversité* que la propriété correspondant à un très improbable adjectif *?inadverse* :

- (4) Et la Suisse [...] a vu [...] la confirmation que l'inadversité dans laquelle elle enfilait ses victoires comme des perles avait duré.

(ii) Les noms de forme *Xisation* dans *Libération* illustreront le second problème. Pour des raisons qui tiennent certainement à ce type textuel de corpus, il apparaît que les noms en *-isation* y sont beaucoup plus fréquents que les verbes en *-is(er)* correspondants, tout particulièrement quand *X* est un anthroponyme ou un toponyme. Il s'ensuit que, si *Xisation* est un hapax, alors *Xis(er)* est assez probablement absent, alors même que le verbe correspondant est possible (cf. tableau 6). Ainsi, il s'agit de déterminer si les noms en *-isation* hapaxiques doivent aussi être pris en considération pour le calcul de *P*, en tant qu'impliquant un verbe en *-is(er)* possible, bien que non utilisé de façon autonome : selon la décision prise, n_1 , donc *P*, pourra varier de façon assez sensible.

Alain Jérôme	alainjéromisation	1	(se) alainjéromiser	0
Bantoustan	bantoustanisation	1	(se) bantoustaniser	0
Belgrade	belgradisation	1	(se) belgradiser	2
Berlusconi	berlusconisation	1	(se) berlusconiser	1
Bonaparte	bonapartisation	1	(se) bonapartiser	0

Tableau 6. Fréquence dans *Libération* de 5 hapax en *Xisation* et des verbes en *Xiser* correspondants où *X* est un anthroponyme ou un toponyme

27. Une autre solution est de considérer qu'il compte pour le calcul de la productivité et de *préf.*, et de *suf.*, en ceci qu'il active chacun de ces procédés dans le lexique mental (cf. Gaeta & Ricca 2003).

3. Conclusion

La recension qui précède a fait apparaître que la principale avancée ces dernières années sur la question de la productivité ne concerne pas tant la définition de la notion elle-même que la proposition de mesures tenant compte de ses facettes qualitative et quantitative. Comme Plag (1999 : 34) en effet, on remarquera que, si la trichotomie disponibilité/rentabilité/régularité opérée par D. Corbin en 1987 a constitué une première étape dans la réflexion sur la notion, les mesures proposées par H. Baayen constituent, elles, une étape décisive en ceci qu'elles encapsulent ces divers paramètres dans des formules mathématiques, permettant ainsi des comparaisons objectives entre procédés. Même si elles posent un certain nombre de problèmes méthodologiques, elles ont le grand mérite d'exister, et de ne pas fonder la notion sur des méthodes plus souvent impressives que réellement scientifiques.

Le second résultat qui se dégage est l'importance accordée, ces dernières années, aux corpus textuels en morphologie constructionnelle. Alors que, jusqu'il y a peu, dans la veine du générativisme, les principaux outils du morphologue étaient les dictionnaires et son intuition (cette bipartition trouve écho dans l'une des acceptions du couple *mots attestés/mots possibles* : cf. § 2.1.1.), on assiste actuellement, dans l'étude du lexique construit, à un déplacement comparable à celui qui s'observe en syntaxe. Désormais, on considère les mots construits *in situ*, dans des contextes produits par des locuteurs non spécialistes du domaine. En procédant de la sorte, on peut dégager les zones actives de la langue et les zones inactives, ce qui évite aux théories de prendre pour principales des données qui sont secondaires.

Références bibliographiques

- ALLEN, Margaret. 1978. *Morphological Investigations*. Ph. D. Diss., University of Connecticut, Storrs, Connecticut.
- ARONOFF, Mark. 1976. *Word Formation in Generative Grammar, Linguistic Inquiry*. Monograph One, Cambridge, Massachusetts/London, England, The MIT Press.
- ARONOFF, Mark & Frank ANSHEN. 1998. « Morphology and the Lexicon: Lexicalization and Productivity ». In A. Spencer & A.M. Zwicky eds.
- ARRIVÉ, Michel, Françoise GADET & Michel GALMICHE. 1986. *La grammaire d'aujourd'hui. Guide alphabétique de linguistique française*. Paris, Flammarion.
- AURNAGUE, Michel & Marc PLENAT. 1997. « Manifestations morphologiques de la relation d'attachement habituel ». In D. Corbin & al. eds, 15-24.
- BAAYEN, Harald. 1992. « Quantitative Aspects of Morphological Productivity ». *Yearbook of Morphology* 1991, 109-149.
- 1993. « On Frequency, Transparency and Productivity ». *Yearbook of Morphology* 1992, 181-208.
- 1994a. « Derivational Productivity and Text Typology ». *Journal of Quantitative Linguistics* 1-1, 16-34.
- 1994b. « Productivity in Language Production ». *LCP* 9, 447-469.
- BAAYEN, Harald & Rochelle LIEBER. 1991. « Productivity and English Derivation: a Corpus-based Study ». *Linguistics* 29-5, 801-843.
- 1997. « Word Frequency Distributions and Lexical Semantics ». *Computers and the Humanities* 30, 281-291.
- BAAYEN, Harald & Antoinette RENOUE. 1996. « Chronicling the Times. Productive Lexical Innovations in an English Newspaper ». *Language* 72, 69-96.
- BAUER, Laurie. 1983. *English Word-formation*. Cambridge, Cambridge University Press.
- 2001. *Morphological Productivity*. Cambridge, Cambridge University Press.

- BLOOMFIELD, Léonard. 1933. *Le langage*. Trad. franç., Paris, Payot, 1970.
- BOOIJ, Geert. 1977. *Dutch Morphology. A Study of Word Formation in Generative Grammar*. Lisse, The Peter de Ridder Press.
- 2002. *The Morphology of Dutch*. Oxford, Oxford University Press.
- BOOIJ, Geert, Christian LEHMAN & Joachim MUGDAN. eds. 2000. *Morphology. An International Handbook on Inflection and Word-Formation*. Vol. 1. Berlin/New York, Walter de Gruyter.
- CHISS, Jean-Louis, Jacques FILLIOLET & Dominique MAINGUENEAU. 2001. *Introduction à la linguistique française. Tome II : syntaxe, communication, poétique*. Paris, Hachette supérieur.
- CORBIN, Danielle. 1976. « Le statut des exceptions dans le lexique ». *Langue française* 30, 90-110.
- 1987. *Morphologie dérivationnelle et structuration du lexique*. 2 vol., Tübingen, Max Niemeyer Verlag ; 2^e éd., Villeneuve d'Ascq, Presses Universitaires de Lille, 1991.
- 1997. « Entre les mots possibles et les mots existants : les unités à faible probabilité d'actualisation ». In D. Corbin & al. eds, 78-89.
- à paraître. *Le lexique construit. Méthodologie d'analyse*.
- CORBIN, Danielle, Bernard FRADIN, Benoît HABERT, Françoise KERLEROUX & Marc PLÉNAT eds. 1997. *Sillexicales 1. Mots possibles et mots existants*, Université de Lille 3.
- CORBIN, Pierre. 1982. « Le monde étrange des dictionnaires (3) : la faisselle et autres contes, scolies sur le changement lexical ». *Linguistiques* 4-1, 125-184.
- CRYSTAL, David. 1992. *An Encyclopaedic Dictionary of Language and Languages*. Oxford, Blackwell Publishers.
- DAL, Georgette. 1997a. « Tous les mots existants sont-ils possibles ? ». In D. Corbin & al. eds, 91-100.
- 1997b. « Du principe d'unicité catégorielle au principe d'unicité sémantique : incidence sur la formalisation du lexique construit morphologiquement ». *Bulag* numéro spécial, 105-115.
- 2002. « À propos d'une idée reçue, ou de la prétendue irrégularité de la dérivation ». *Bulag* 27, 57-73.
- DIEZ, Friedrich. 1838. *Grammatik der romanischen Sprachen II*. Bonn, Weber.
- DI SCIULLO, Anna Maria & Edwin WILLIAMS. 1987. *On the Definition of Word*. Cambridge, Mass., The MIT Press.
- DRESSLER, Wolfgang U. & Mária LADÁNYI. 2000. « Productivity in Word Formation (WF): a Morphological Approach ». *Acta Linguistica Hungarica* 47-2, 103-144.
- DUBOIS, Jean, Mathée GIACOMO, Louis GUESPIN, Christiane MARCELLESI, Jean-Baptiste MARCELLESI & Jean-Pierre MÉVEL. 1999. *Dictionnaire de linguistique et des sciences du langage*. Paris, Larousse-Bordas/Her.
- DUGAS, André & Christian MOLINIER eds. 1992. *Langue Française* 96. *La productivité lexicale*.
- FRADIN, Bernard. 1997. « Les mots-valises : une forme productive d'existants impossibles ? ». In D. Corbin & al. eds, 101-110.
- 2003. *Nouvelles approches en morphologie*. PUF, Paris.
- FRAUENFELDER, Uli & Robert SCHREUDER. 1992. « Constraining Psycholinguistic Models of Morphological Processing and Representation: the Role of Productivity ». *Yearbook of Morphology* 1991, 165-183.
- GAETA, Livio & Davide RICCA. 2003. « Italian Prefixes and Productivity: a Quantitative Approach ». *Acta Linguistica Hungarica* 50, 89-108.
- GUILBERT, Louis. 1975. *La créativité lexicale*. Paris, Larousse.
- HALLE, Morris. 1973. « Prolegomena to a Theory of Word Formation ». *Linguistic Inquiry* 4/1, 3-16.
- HAY, Jennifer. 2001. « Lexical Frequency in Morphology: Is Everything Relative? ». *Linguistics* 39-6, 1041-1070.
- HAY, Jennifer & Harald BAAYEN. 2002. « Parsing and Productivity ». *Yearbook of Morphology* 2001, 203-235.
- HUOT, Hélène. 1997. « Des mots possibles aux mots existants : système morphologique et structuration du lexique ». In D. Corbin & al. eds, 141-150.
- JESPERSEN, Otto. 1942. *A Modern English Grammar on Historical Principles*. Vol. VI : *Morphology*. Copenhagen, Munksgaard.
- KASTOVSKY, Dieter. 1986. « The Problem of Productivity in Word Formation ». *Linguistics* 24-3, 585-600.
- KILFER, Ferenc. 2000. [article n° 32 : Regularity]. In G. Booij & al. eds, 296-302.
- KIPARSKY, Paul. 1982. « Lexical Morphology and Phonology ». In I.S. Yang ed., *Linguistics in the Morning Calm*, Seoul, Hanshin Publishing Company, 1-91.
- KOEFOD, Geert & Jaan VAN MARIE. 2000. [article n° 33 : Productivity]. In G. Booij & al. eds, 303-311.

- KROTT, Andrea, Robert SCHREUDER & Harald BAAYEN. 1999. « Complex Words in Complex Words ». *Linguistics* 37-5, 905-926.
- LEHMANN, Alise & Françoise MARTIN-BERTHELET. 1998. *Introduction à la lexicologie. Sémantique et morphologie*. Paris, Dunod.
- Libération. Les archives de janvier 1995 à décembre 1999*. CEDROM-SNi inc. (Montréal), 1999.
- LIEBER, Rochelle. 1992. *Deconstructing Morphology. Word Formation in Syntactic Theory*. Chicago and London, The University of Chicago Press.
- LM = *Le Monde sur CD-ROM*. SA Le Monde (Paris) – CEDROM-SNi inc. (Montréal), 1999.
- LMD = *Le Monde Diplomatique 1984-1998*. CEDROM-SNi inc. (Montréal), 1998.
- LYONS, John. 1977. *Semantics*. Cambridge, Cambridge University Press.
- MARCHAND, Hans. 1951. « Phonology, Morphonology and Word-formation ». *Neuphilologische Mitteilungen* 52, 87-95.
- MARLE, Jaan van. 1992. « The relationship between morphological productivity and frequency: a comment on Baayen's performance oriented conception of morphological productivity ». *Yearbook of Morphology* 1991, 151-163.
- MATTHEWS, Peters H. 1991. *Morphology*. Cambridge, Cambridge University Press, second edition; first ed. 1974.
- MOUNIN, Georges. éd. 1974. *Dictionnaire de la linguistique*. Paris, P.U.F.
- PICHON, Édouard. 1942. *L'enrichissement lexical dans le français d'aujourd'hui. Les principes de la suffixation en français*. Paris, Éditions d'Artrey.
- PLAG, Ingo. 1999. *Morphological Productivity. Structural Constraints in English Derivation*. Berlin/ New York, Mouton de Gruyter.
- POITOU, Jacques. 1997. « Régularité, analyses quantitatives et productivité : quelques remarques ». In D. Corbin & al. eds, 221-230.
- PRE = *Version électronique du Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française, Nouvelle édition (version 2)*. Dictionnaires Le Robert/VUEF, 2001.
- RAINER, Franz. 1987. « Produktivitätsbegriffe in der Wortbildungstheorie ». In W. Dietrich, H.M. Gauger & H. Geckeler eds, *Grammatik und Wortbildung romanischer Sprachen*, Tübingen, Narr, 187-202.
- 2000. [article n° 85 : Produktivitätsbeschränkungen]. In G. Booij & al. eds, 877-885.
- SANTEN, Ariane van. 1992. *Productiviteit in taal en taalgebruik*. Unpublished doctoral dissertation, University of Leiden.
- SCALISE, Scalise. 1984. *Generative Morphology*. Dordrecht (Holland)/Cinnaminson (U.S.A.), Foris Publications.
- SCHULTINK, Henk. 1961. « Produktiviteit als Morfologisch Fenomeen ». *Forum der Letteren* 2, 110-125.
- 1992. « Herkomst en groei van het begrip 'productiviteit' als morfologisch fenomeen, Produktiviteit als morfologisch fenomeen ». In M. Klein ed., *Nieuwe eskapades in de neerlandistiek*, Groningen, Wolfers-Noordhoff, 187-201.
- SPENCER, Andrew & Arnold M. ZWICKY eds. 1998. *The Handbook of Morphology*. Oxford, Blackwell Publishers.
- TLF = *Trésor de la langue française. Dictionnaire de la langue du XIX^e et du XX^e siècle (1789-1960)*. 16 vol., Paris, Éditions du CNRS (t. 1-10) / Gallimard (depuis le t. 11), 1971-1994.
- TRASK, Robert Lawrence. 1993. *A Dictionary of Grammatical Terms in Linguistics*. London, Routledge.
- ZWANENBURG, Wiecher. 1983. *Productivité morphologique et emprunt. Étude des dérivés déverbaux savants en français moderne*. Amsterdam/Philadelphia, John Benjamins Publishing Company.
- ZIMMER, Karl E. 1964. *Affixal Negation in English and Other Languages: an Investigation of Restricted Productivity*. Supplement to *Word* 20-2, Monograph 5.