

Bernard Fradin, UMR 7110, GDR 2220 CNRS & Paris 7
bernard.fradin@linguist.jussieu.fr
Nabil Hathout, UMR 5610, GDR 2220 CNRS & U. Toulouse le Mirail
nabil.hathout@univ-tlse2.fr
Fanny Meunier, UMR 5596, CNRS & U. Lyon 2

LA SUFFIXATION EN *-ET* ET LA QUESTION DE LA PRODUCTIVITÉ

1. Introduction

L'étude des mots suffixés en *-ET* du point de vue de la productivité se justifie par le statut ambivalent que présente ce suffixe sous ce rapport. D'une part, cette suffixation est un procédé apparemment très productif, ce qui se manifeste par la création de nombreux types nouveaux (*torsette, aspirette, scoopinet, cadrette*) ; d'autre part, un nombre considérable de suffixés en *-ET* sont opaques et ne se rattachent à aucun patron de construction actuellement disponible (voir §4). Le premier constat nous a incités à entreprendre de mesurer la productivité de *-ET* sur corpus, ce qui a débouché sur la présente étude. Le second nous a imposé de trier soigneusement les formes en *-ET* qu'on doit prendre en compte et nous a obligé à élaborer des critères permettant de discriminer les formes opaques des transparentes, qui seules sont pertinentes pour mesurer la productivité (Van Marle 1985).

Il est notoire que les mots suffixés en *-ET* présentent une palette de significations très grande (voir §2), qui toutes relèvent d'une règle différente, appelée ici patron. Dans la mesure où il n'y a pas de raison a priori pour que la productivité soit identique pour chaque patron, nous n'avons pas cherché à mesurer la productivité globale de *-ET* et avons opté pour une mesure différenciée pour chacun. Ceci suppose qu'on détermine du mieux possible les patrons construisant chaque signification. Cette exigence nous a conduit à adopter une démarche ascendante : nous sommes partis des sens identifiables pour chaque forme dans les textes du corpus pour aboutir, par classements et regroupements successifs, à une classification englobante hiérarchisée.

Pour ce qui regarde la sémantique, nous avons fait comme si nous avions affaire à un seul suffixe *-ET*. En revanche, du point de vue de la forme nous avons considéré *-et* et *-ette* comme deux suffixes différents et les avons traités séparément dans nos décomptes. Méthodologiquement, ces choix étaient les seuls envisageables, puisqu'on voulait mesurer la productivité relative de chaque forme. De fait, il s'avère que *-ette* est plus productif quantitativement que *-et* (§3).

La deuxième section donne une présentation très résumée du système de la suffixation en *-ET* en français (voir (Fradin 2003) pour le détail). La troisième expose la méthode utilisée pour préparer le corpus et mesurer la productivité. La quatrième donne les principaux résultats de mesure de la productivité que nous avons obtenus et les commente.

2. Le système de la suffixation en *-ET*

2.1. La morphologie diminutive

La suffixation en *-ET* est le principal procédé de formation des diminutifs en français contemporain. Ce procédé relève de la morphologie évaluative. A la différence de la morphologie ordinaire, qualifiée parfois de conceptuelle parce qu'elle fournit des moyens permettant de catégoriser les objets dont on parle, la morphologie évaluative exprime une appréciation sur ceux-ci et renseigne sur le positionnement du locuteur vis-à-vis de ce qu'il décrit ou de son interlocuteur. De ce fait, elle apporte très souvent des connotations pragmatiques.

Notre objectif est d'esquisser comment s'organise une sous-partie du système des diminutifs en français. Notre description se limite à la suffixation en *-ET* qui construit des noms. Nous laissons de côté celle qui dérive des adjectifs (*gentillet, tristounet*) ou des verbes (*voleter, becqueter*), les procédés suffixaux plus anciens, la plupart du temps improductifs (*diable > diablotin, lion > lionceau, ours > ourson, balle > ballot*, etc.). Nous ne dirons rien non plus des aspects morphophonologiques de cette suffixation et notamment de l'intercalation fréquente d'un segment phonique (*bras > brac-el-et*) et renvoyons à (Plénat & Roché 2001) où ce phénomène reçoit un traitement nouveau.

L'analyse présentée ici a largement tiré parti de (Dal 1997) et (Delhay 1996), deux travaux récents importants sur la morphologie diminutive du français, qui intègrent les acquis des études antérieures.

2.2. L'organisation sémantique des dérivés en *-ET*

Les significations considérées dans ce travail reprennent celles qui ont été distinguées dans la littérature ou les dictionnaires. A été comptée comme signification distincte toute interprétation induite par l'emploi d'un lexème construit en *-ET* dans un énoncé, qui était : attestée, reproductible, non inférable d'une signification déjà répertoriée¹. Les écarts à ce dernier principe seront mentionnés et justifiés.

2.2.1. Bipolarité

Nous faisons l'hypothèse que l'organisation en question est bipolaire en français (fig. 1 et 2). Le premier pôle est nommé « pôle Référent » parce qu'il regroupe des sens ayant trait aux propriétés constitutives ou fonctionnelles de l'entité dénotée par le dérivé en *-ET*. Toutes les significations regroupées dans ce pôle mettent en jeu la notion d'échelle. Il peut s'agir d'une échelle graduée ou d'une échelle ne présentant que deux valeurs antonymiques e.g. [+fem] vs [-fem]. L'effet sémantique induit par la suffixation en *-ET* pour ce pôle est de l'ordre de la diminution : en gros, l'entité dénotée par le dérivé possède une propriété à un degré inférieur au degré standard fixé, en l'occurrence, par le référent du lexème-base². Le second pôle regroupe toutes les significations ayant trait à l'interaction du locuteur avec le référent. Alors que le premier pôle se situe dans une

¹ L'interprétation 'lieu planté de X' associée à *olivette, aulnette, coudrette* par des dictionnaires n'a pas été retenue car (a) nous n'avions pas d'attestation, (b) le suffixe y vient du latin *-etum* (Dal 1997 : 80). *Farigoulette* vient du provençal. La discussion sur le placement des diverses significations les unes par rapport aux autres n'est pas close.

² Pour les échelles binaires, on a affaire à une inversion de polarité e.g. *merle / merlette*.

perspective descriptive, le second se situe dans une perspective qualitative³. Ce pôle englobe tous les cas où « le dérivé ne renseigne pas sur une propriété du référent, mais sur la relation que pose l'énonciateur entre lui-même et le référent : ce qui est diminué, c'est la distance entre le locuteur et ce dont il parle. On peut parler alors d'une forme d'appropriation du monde par le locuteur. » (Delhay 1996: 191). A la différence du premier, ce pôle ne fait pas intervenir d'échelle. Un troisième pôle est envisageable, qui regrouperait les significations relatives à l'interaction du locuteur avec l'interlocuteur (pôle Interlocuteur), interaction qui vise à l'intégrer dans sa sphère ou à se faire admettre dans la sienne. Ce pôle comprendrait les diminutifs dont l'emploi manifeste l'empathie du locuteur vis-à-vis de son interlocuteur : *diminutiva puerilia*, atténuation de l'acte de langage, etc. (Dressler & Merlini Barbaresi 1994). Ce pôle n'existe pas en français, ce qui explique qu'on n'ait aucun de ces emplois pourtant caractéristiques de beaucoup de langues à morphologie évaluative développée⁴ (voir (Fradin 1999) pour une discussion).

2.2.2. Approche en MCI

Notre seconde hypothèse est que les relations entre les diverses significations regroupées sous chaque pôle ainsi que les liens entre les deux pôles peuvent être décrits au moyen d'une catégorie radiale organisée en Modèle Cognitif Idéalisé (MCI) (Lakoff & Johnson 1980). En général, les MCI sont utilisés pour rendre compte du réseau de significations associé à un lexème donné (polysémie). Dans le cas présent, et en simplifiant un peu, les nœuds du réseau ne sont pas des significations mais les règles morphologiques qui permettent de construire les dérivés en *-ET*. Ces règles, appelées aussi Règles Morphologiques Lexicales (RML) à la suite de (Stump 1991) correspondent aux feuilles dans la représentation du réseau donné en figure 1 et 2. Une justification de l'utilisation des MCI est donnée dans (Fradin 2003) auquel nous renvoyons. Notons que (Jurafsky 1996) propose un modèle de type MCI pour rendre compte des valeurs sémantiques universellement observables de la catégorie du diminutif et de son évolution diachronique .

Les mécanismes morphologiques

Notre conception de la dérivation diminutive repose sur l'hypothèse que l'affixe *n'y* est pas nécessairement un opérateur s'appliquant sur la signification du lexème base pour donner la signification construite. Conséquemment, il n'est pas exigé que la signification du dérivé soit entièrement calculable à partir de celle de la base pour qu'il y ait dérivation. Un tel point de vue est soutenable parce que nous mettons l'accent sur l'output : la signification du dérivé en *-ET* est envisagée du point de vue de

³ L'évaluation associée à la morphologie évaluative « peut faire référence à des propriétés 'physiques', objectives et vérifiables des entités considérées ou bien aux qualités de ces même entités telles qu'elles sont subjectivement perçues par les locuteurs » (Grandi 2002 : 31).

⁴ Deux exemples pour fixer les idées : (a) et (c) sont normaux en italien et en allemand mais leurs équivalents impossibles en français : (a) *Anch'io avrei una domand-ina* (Dressler & Merlini Barbaresi 1994 : 333) (b) **Moi aussi, j'aurais une questionn-ette*. (c) *Ah, was für ein köstliches Supp-erl!* (d) **Ah! quelle soup-ette délicieuse!* (ibid. : 89).

ce que dénote ce dérivé. L'idée est que ces dérivés sont créés avant tout pour fournir des dénominations (point de vue onomasiologique), l'entité à dénommer étant déjà identifiée dans le monde.

La dérivation diminutive en *-ET* peut se faire selon deux régimes. Dans le premier, qui concerne le pôle référent exclusivement, le dérivé est construit classiquement au moyen d'une Règle Morphologique Lexicale qui, à la fois, spécifie l'apport sémantique du procédé (cf. les distinctions R1-R10 §2.4) et établit la relation entre le lexème-base et le lexème-dérivé. Dans le second, qui concerne le pôle référent, l'effet sémantique du procédé est découplé de la mise en relation de la base et du dérivé. L'apport sémantique associé à *-ET* dépend uniquement du type sémantique du référent du N dérivé. Par exemple, un nom comme *camionnette* héritera de la propriété générale 'commode à manœuvrer' sur la base du fait que ce nom dénote une entité relevant de la classe des 'moyens de transports'⁵. Cette information apparaîtra chaque fois que le N dérivé dénotera une entité relevant de cette classe. Dans le cas où le dérivé dénote un outil ou un ustensile (*carrelette*, *passette*, *moulinette*, *calcullette*, *arielette*), la suffixation en *-ette* apportera une information du type 'facile à manier/à utiliser'. Il y a donc un conditionnement réciproque entre l'apport d'information associé à la suffixation et le type sémantique du référent.

Dans le mécanisme qu'on vient de décrire, l'interprétation du dérivé n'est pas construite à partir de celle du lexème-base, puisque seul le référent du dérivé compte. Ce mécanisme ne peut donc établir le lien entre base et dérivé, quand il existe. Nous faisons l'hypothèse qu'un autre mécanisme le prend le charge, qui est basé sur des inférence lexicales très générales (ILG, Inférence Lexicale Générale). Ce mécanisme vise à combiner de manière cohérente et optimale des informations contenues dans la représentation sémantique (RS) du lexème-dérivé, ou inférables à partir de son type sémantique, et des informations provenant de la RS du lexème-base. Le mécanisme en question étant décrit en détail dans (Fradin 2003), nous ne ferons que l'illustrer ici au moyen de quelques exemples. Prenons *bleu* et *bleuet*. Le premier est un adjectif dénotant une couleur (i.e. un objet de type 'couleur') et le second un N dénotant une fleur des champ dont la couleur est effectivement bleue. Le lien entre les deux est trivial: en tant que fleur, le bleuet hérite de la propriété d'avoir une couleur (mais aussi une odeur, une forme, etc.). Le prédicat *bleu*' instancie la valeur de la propriété couleur pour la fleur qu'est le bleuet. Le deuxième exemple sera celui de *calcullette*. En tant qu'il dénote un artefact fonctionnel, ce N hérite des propriétés attachées à ce type. Une de celles-ci dit qu'un artefact fonctionnel est un objet avec lequel un agent fait quelque chose, ce qui peut se noter comme (i) ($\lambda x_3. V' \bullet e \bullet x_1 \bullet x_2 \bullet (\text{au-moyen-de}' \bullet x_3)$). La relation entre *calcullette* et le verbe *calculer* repose sur le fait que ce dernier instancie le verbe agentif V' en (i), ce qui apparaît à la lecture de la RS du dérivé, qui serait quelque chose comme (1):

⁵ La classe des moyens de transports sera notée Lv (véhicule). Bien qu'il soit énoncé sous une forme absolue, le contenu de l'inférence ne prend de valeur que par comparaison avec d'autres types de véhicules. Les expressions relevant des sous-groupes Lh (habit) et Lf (fourrure) n'ont pas été distinguées dans la base de données. Dans les décomptes, elles sont comptabilisées sous Lg (général).

$$(1) \quad (\lambda_{x_3}. \text{calculer}' \bullet e_1 \bullet x_1 \bullet x_2 \bullet (\text{au-moyen-de} \bullet x_3) \wedge \text{outil}' \bullet x_3 \wedge \text{humain}' \bullet x_1 \wedge \text{facile}' \bullet e_1)$$

N'étant pas déterminées par le processus qui construit la signification, comme c'est le cas avec les RML, les manières d'établir une relation sémantique entre la base et le dérivé présentent quelques variations mineures.

Nous allons maintenant présenter une à une les significations discriminées. Les figures 1 et 2 donnent une vue partielle de l'organisation du réseau de significations associées à *-ET*.

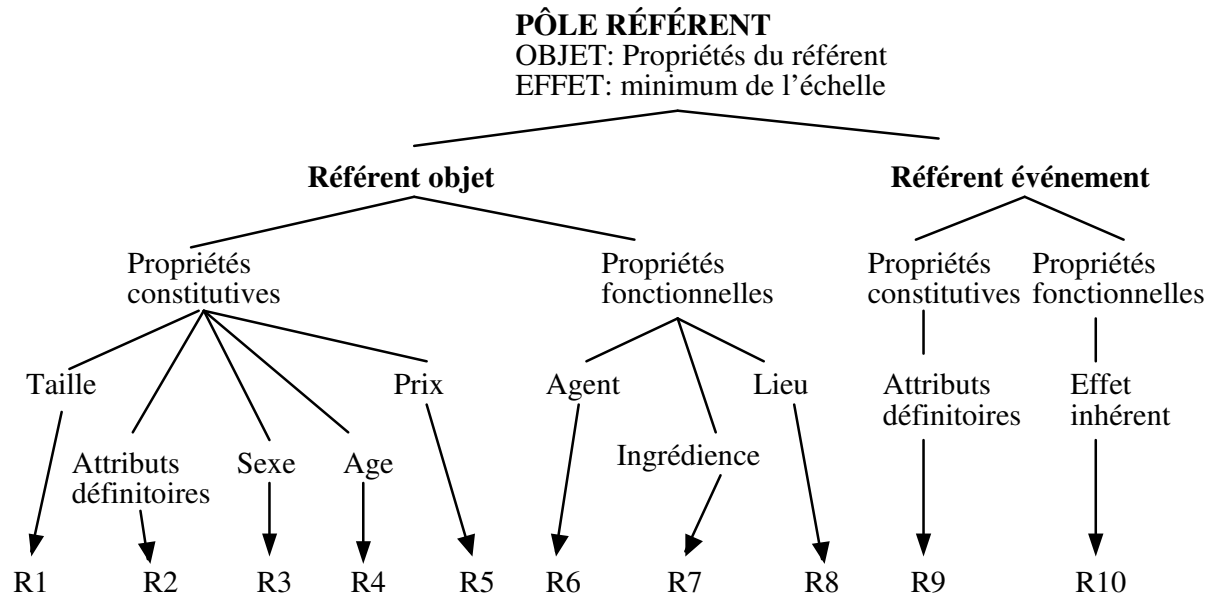


Fig. 1

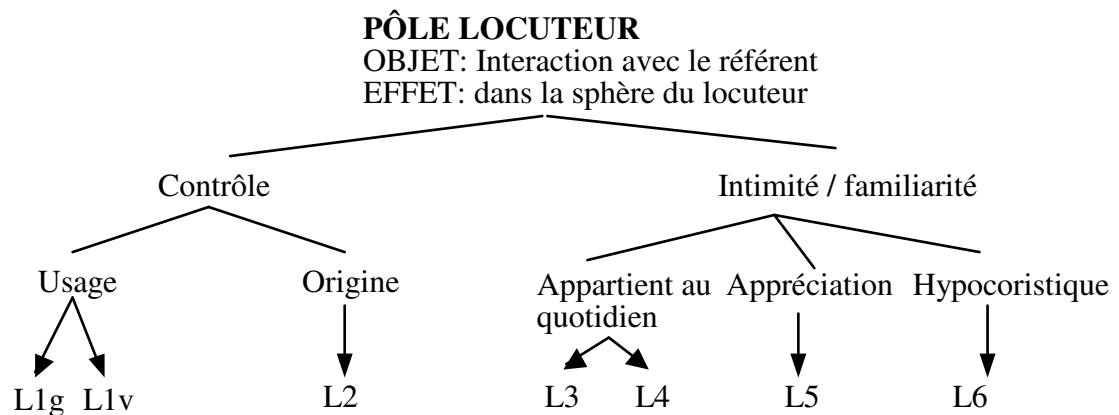


Fig. 2

2.4. Le pôle référent

Le référent peut être un objet (animé ou non) dénombrable (*brebiette*, *clochette*) ou bien un événement (*causette*). La propriété faisant l'objet de l'évaluation peut être une propriété constitutive de l'objet ou bien une propriété fonctionnelle. Toutes les propriétés constitutives, sauf en R3

(*merle/merlette*), mettent en jeu la notion d'échelle. Dans tous ces cas, la suffixation en *-ET* indique que le degré d'une dimension DMS_i que possède le référent du lexème-dérivé est inférieur au degré que présente l'objet de référence (odr), en l'occurrence le référent type du N-base (objet ou événement). Ce référent est sémantiquement du type objet, à l'exception de R8 où il est un lieu et de R9 où il est un événement. Le changement sémantique est noté dans le tableau 2 pour chacune des significations du pôle Référent⁶. Par exemple, dans le cas de R1, *clochette* dénote un objet qui a une taille inférieure à celle d'une cloche, laquelle est censée avoir la taille standard pour les cloches. Autres exemples de R1 : *balisette, brochette, brochurette, châtelet, maisonnette...*

En R2 plusieurs propriétés du référent sont prises en compte simultanément. Les cas tournent tous autour de l'idée que le référent du dérivé présente sous une forme édulcorée les (ou certaines des) propriétés caractéristiques du référent du lexème-base : soit ces propriétés ne se retrouvent pas dans le référent du dérivé, soit elles s'y retrouvent sur un mode atténué e.g. *histoire / historiette, amour / amourette, opéra / opérette, cigare / cigarette, berline / berlinette, sucre / sucrlette*. Dans le tableau 2, la formule vise à dire que pour toute propriété P_i que possède le référent du N-base ($P_i \cdot x_1$) et pour toute dimension de cette propriété, soit le référent du dérivé ne possède pas cette propriété, soit la valeur de sa dimension est moindre. En toute logique, le cas R1 est un sous-cas de R2. S'il ne figure pas sous ce dernier, c'est à cause du nombre des types qu'il recouvre et de son importance comme prototype dans le dispositif cognitif de l'évaluation.

En R3 la dimension concerne exclusivement la sexuisemblance. R3 regroupe les significations où l'on bascule d'une valeur à l'autre⁷ : mâle vs femelle e.g. *fliquette, punkette, fantômette, beurette, PDGette, cadrette*. Ces dérivés présentent fréquemment une connotation hypocoristique (L6).

En R4 la propriété en question est l'âge. L'interprétation est du type 'jeune de l'espèce Nb' e.g. *poulet, porcelet*. Dans la mesure où il existe une implication générale : 'jeune de l'espèce X' → 'plus petit en taille que l'adulte de X', ce cas pourrait être vu comme une extension de R1.

La dimension retenue ici pour R5 est le prix. Plusieurs de ces dérivés dénotent des imitations de fourrure dont la valeur est bien moindre que la fourrure authentique : *chinchillette, castorette, herminette₂, angorette*. On pourrait aussi choisir de faire de cette signification un sous-type de R2 dans la mesure où les référents de R4 sont des ersatz du référent du N-base. La seule différence est que dans ce dernier cas, le côté ersatz est sciemment revendiqué parce qu'il rend ces produits plus accessibles aux chalands. Pour cette raison, on retrouve les dérivés de type R4 dans le pôle Locuteur (L1).

R6 regroupe des dérivés dont le nom-base dénote des individus ou des entités qui exercent une fonction ou sont dotés de pouvoirs. La suffixation *-ET* indique que ces pouvoirs ou ces fonctions sont amoindris, qu'ils n'ont plus les effets escomptés. *Roitelet, foutriquet, bombinette* sont de ce type.

⁶ Ces notations sont une sténographie simplifiée de ce qui se passe et nullement une représentation formelle en bonne et due forme. Les propriétés constitutives pour les dimensions concernées s'expriment par des adjectifs (*grand, cher, vieux*, etc.) dont le degré standard auquel ils sont vrais est fixé en contexte (Kennedy & McNally 1999).

⁷ Inversement, le suffixe *-et* a servi à marquer des basculements de femelle vers mâle e.g. *lanier / laneret* cf. (Dal 1997 : 23-24).

La fonction discriminée par R7 est celle d'une substance servant d'ingrédient non majoritaire dans une préparation culinaire e.g. *anisette*, *vinaigrette*, *citronnette* (Dal 1997: 142). Le français dénomme les préparations où le N-base dénote l'ingrédient majoritaire par des dérivés en *-ade* e.g. *poivrade* (Aliquot-Suengas 1997) ; d'où les paires minimales *citronnette / citronnade*, *anisette / anisade*.

R8 rassemble les dérivés qui dénotent des lieux spécialement destinés à effectuer l'activité exprimée par le verbe-base. Le suffixe *-ET* indique que celle-ci est temporaire. Il y a quatre N de ce type : *buvette*, *couchette*, *cachette*, *oubliette* (lieu où l'on oublie exprès quelqu'un)⁸. Le caractère temporaire (vs. permanent) est net dans les paires *cachette / cache* et *couchette / couche*.

Le type R9 est le pendant de R2 pour les événements : ceux-ci présentent une ou plusieurs dimensions (durée, amplitude, intensité) diminuées par rapport à celles que manifeste le verbe base. Ces dérivés sont assez nombreux : *giclette*, *sifflette*, *reniflette*, *risette*, *grimpette*, *causette*, *tapette*, *trempelette*, *bronzette*. Ce type aussi donne lieu à des paires minimales : *causerie / causette*, *tape / tapette*, *giclée / giclette*, *bronzage / bronzette*.

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

Le tableau 2 résume l'apport sémantique de la suffixation en *-ET* pour le pôle Référent.

	EXEMPLE	ENTREE	SORTIE
R1	<i>clochette</i>	objet'•x ₁ ∧ nombrable'•x ₁	inf•(deg•x ₂ •taille')•(deg•x ₁ •taille')
R2	<i>opérette</i> , <i>cigarette</i>	entité'•x ₁	∀P ₁ . P ₁ •x ₁ , ∧ ∀DMS _k ∈ P ₁ , (P ₁ •x ₂ => inf•(deg•x ₂ •DMS _k)•(deg•x ₁ •DMS _k)) ∨ ¬P ₁ •x ₂
R3	<i>merlette</i>	mâle'•x ₁	femelle•x ₂
R4	<i>poulet</i>	animal'•x ₁	inf•(deg•x ₂ •âge')•(deg•x ₁ •âge')
R5	<i>castorette</i>	fourrure'•x	inf•(deg•x ₂ •valeur')•(deg•x ₁ •valeur')
R6	<i>roitelet</i>	exercer'•x ₁ •pouvoir' => (R ₁ ...R _n)	exercer'•x ₂ •pouvoir' => ¬(R ₁ ...R _n)
R7	<i>vinaigrette</i>	substance'•x ₁	substance'•x ₁ •(dans•x ₂) ∧ préparation'•x ₂ ∧ ¬majoritaire'•x ₁ •(dans•x ₂)
R8	<i>couchette</i>	V•e•x ₁ ...	V•e•x ₁ •(dans•x ₂) ∧ temporaire'•e
R9	<i>giclette</i> , <i>causette</i>	V•e ₁ •x...	V•e ₂ •x... ∧ ∀DMS _k , inf•(deg•e ₂ •DMS _k)•(deg•e ₁ •DMS _k)
R10	<i>réformette</i>	V•e•x ₁ ∧ effet-de'•e ₁ •e ₂	V'•e•x ₃ ∧ effet-de'•e ₃ •e ₄ ∧ ∀DMS _k , inf•(deg•e ₄ •DMS _k) •(deg•e ₂ •DMS _k)

Tableau 1. Sémantique du pôle référent

2.5. Pôle Locuteur

⁸ Ce dérivé se retrouve aussi dans le pôle Locuteur sous L3.

Les deux sous-pôles ‘Contrôle’ et ‘Intimité / familiarité’ correspondent, en gros, à la distinction entre ce sur quoi le locuteur peut agir et ce sur quoi il n’a pas de prise. ‘Contrôle’ rassemble exclusivement les dérivés dont le référent est un artefact fonctionnel concret (instrument, outil, arme, pièce de vêtement, etc.) ou abstrait (mesure politique ou administrative). Le second sous-pôle rassemble au contraire des dérivés dénotant en majorité des espèces naturelles (L3) ou des animés (L5). Les dérivés sous L4 dénotent généralement des aliments. Leur présence à cette place sera motivée plus loin.

Le groupe L1g (L1 général) regroupe les dérivés dénotant des objets conçus pour servir à l’accomplissement d’une fonction précise (artefacts fonctionnels). Pour cette raison, beaucoup dénotent des outils ou des ustensiles : *allumette, baïonnette, boulet, serpette, mesurette, raclette, fourchette, douchette, lunette, mesurette, pipette, mitraillette, mouillette, moulinette, sonnette, sucette, traceret, arielette*, etc. L’idée, souvent notée à leur propos, est qu’ils sont des objets maniables, qui facilitent l’exécution de certaines tâches et dont il est commode de se servir. Pour cette raison, la sémantique qui sera associée à L1 exprime la maniabilité, la facilité d’usage. Elle est génériquement notée ici ‘maniable’*x’. A l’intérieur de L1, on peut, si l’on veut, distinguer des sous-groupes pour lesquels le sens qu’apporte la suffixation varie en fonction du type de l’objet dénoté. Pour les véhicules au sens large (*camionnette, fourgonnette, patinette, avionnette, caninette*), il serait ‘facile à conduire / manœuvrer’ ; pour les parties de vêtement (*épaulette, talonnette, cuissette, manchette, poignet*), il serait quelque chose comme ‘améliore (la fonction d’)un vêtement’. Pour les fourrures (*chinchillette*, etc.), déjà vues au pôle Référent en tant qu’imitations, il pourrait être ‘(plus) faciles à porter’ pour signifier qu’elles peuvent plus facilement être achetées et sont moins fragiles que les vraies fourrures.

En L2, le contrôle se traduit par le fait que le référent dénoté par le dérivé (une mesure administrative) est quelque chose qui n’aurait pas existé sans l’action du référent du N-base (une personnalité politique). Il existe une filiation métaphorique entre la mesure en question et celui qui l’initie, ce qui rattache L2 au cœur du MCI des diminutifs. Les expressions de L2 sont à ranger sous le pôle Locuteur parce qu’on ne retrouve dans aucune l’effet de diminution caractéristique du pôle Référent. L’effet interprétatif associé à L2 est plutôt celui d’une mesure administrative qui apporte un bénéfice inattendu, mais modeste, dans le quotidien. Ceci apparente sans ambiguïté L2 à L5 et L6. Le procédé L2 est productif, comme la série *chalandonnette, baladurette, périssolette, arthuisette, aubryette, jospinette* en témoigne.

Les dérivés L3 nous placent au cœur de la valeur d’appropriation reconnue à la suffixation *-ET*. Celle-ci exprime l’acointance qui existe entre le locuteur et le référent du dérivé. Elle se traduit très souvent par la manière même dont le dérivé est construit : la base exprime une caractéristique physique ou comportementale saillante du référent du dérivé, caractéristique qui n’est perceptible qu’à celui qui est familier du référent, qui le cotoie régulièrement. Les caractéristiques suivantes sont les plus communes :

- couleur : *roussette, rouget, bleuet, jaunet* (plante aquatique), *fauvette*.
- goût : *farinet, doucette* (= mâche).
- odeur : *fenouillette* (type de pomme), *muguet* (< *musc*).
- forme : *rosette, sapinette* (type d’épicéa), *pommette*.

- nourriture : *chardonneret* (cet oiseau est friand de chardons).
- habitat : *bergeronnette*, *souchette* (champignon qui croît sur les souches).
- saison d'apparition : *pâquerette* (fleurit vers Pâques).

La seule différence entre les dérivés L3' et ceux de L3 tient au fait que les premiers dénotent des espèces nuisibles ou mauvaises. Dénommer celles-ci au moyen du diminutif, c'est faire comme si elles appartenaient à la sphère familière du locuteur alors même qu'il a toutes les raisons de s'en méfier. D'où la lecture propitiatoire associée à ces dérivés. L'exemple canonique est celui de *belette* (< *belle*). Nous avons considéré que *riflette* 'guerre', qui apparaît dans le corpus dans *être bon pour la riflette*, est une réactivation de ce schéma. *Oubliette* s'y rattacherait aussi (bien que ce soit un artefact).

Les dérivés L4 ont affaire au goût et à l'alimentation. Certains sont anciens comme *andouillette*, *rillettes*, *tartiflette*. D'autres sont récents et dénotent des aliments d'une variété nouvelle (*gariguettes*, *torsadette*, *macaronette*) ou d'un apprêt nouveau et soigné (*croquette*, *goujonnette*, *gigolette*, *jambonnette*). L'emploi du diminutif vise à induire que ces produits relèvent de la sphère des aliments proches du goût des locuteurs, ou qu'ils ont été préparés pour leur plaisir (aliments en portions). Que ces dénominations relèvent d'une démarche commerciale n'enlève rien au fait que c'est la suffixation en *-ET* qu'on a choisie. Elles jouent sur le fait que le type d'alimentation, et a fortiori la cuisine, fonctionne comme un identifiant culturel et social très fort (Bourdieu 1979). Nous proposons que l'effet de sens générique associé au type L4 énonce que le goût du référent est apprécié par le locuteur.

Les dérivés L5 sont des hypocoristiques. Il s'agit d'une fonction ancienne de la suffixation en *-ET* qui est toujours productive e.g. *cochonnette*, *pépette*, *pulpinette*. L'examen des corpus indique que les suffixes élargis par *-oun-*, et parfois *-in-*, tendent à s'employer pour marquer la valeur hypocoristique pure e.g. *bichounette*, *sitounet*, *pitchounette*, *scoopinet*, *zigounette*. De cette façon, on évite la confusion avec les dérivés de type R1 e.g. *bichette* ou R3 *cadrette*, *vampirette*, *punkette*.

Le tableau 3 résume les instructions sémantiques pour le pôle L. L'information mentionnée dans la colonne Déclencheur donne l'information de la RS du lexème dérivé qui détermine quelle est l'information apportée par la suffixation.

	EXEMPLE	DECLENCHEUR	APPORT SEMANTIQUE
L1g	<i>serpette</i> ,	objet-fonctionnel'•x	maniable'•x
L1v	<i>camionnette</i>	véhicule'•x	manoeuvrable'•x
L1f	<i>chinchillette</i>	imitation-fourrure'•x	abordable'•x
L1h	<i>talonnette</i>	partie -vêtement'•x	améliorer'•e•x•y \wedge vêtement'•y
L2	<i>baladurette</i>	mesure-adm'•x	favoriser'•e•x•y \wedge mesure-adm•x
L3	<i>bergeronnette</i>	espèce-nat'•x	dans•x•y \wedge sphère-de-loc •y
L4	<i>rillettes</i>	aliment'•y	goût'•z•y \wedge apprécier'•e•x•z \wedge loc'•x
L5	<i>fripounette</i>	personne'•x	avoir-affection-pour'•e•x•y \wedge loc'•x

Tableau 2. Sémantique du pôle locuteur

2.6. Transparence et opacité

Sont considérés comme transparent les mots en *-ET* qui (i) sont analysables en synchronie et (ii) pour lesquels il existe une RML disponible. « Analysable » veut dire ici segmentable en une partie suffixale dont la phonologie est /Et/ ou /E/ et une partie radicale X qui doit correspondre à la rubrique phonologique (base ou base supplétive) d'un lexème figurant dans le lexique. Une RML est une correspondance réglée entre un lexème-base et un lexème-dérivé. Les RML n'existent que pour les dérivés du pôle Référent, si l'on excepte L2.

Quatre degrés d'opacité sont distinguables. Pour le premier le lexème est segmentable en synchronie mais aucune RML ayant cours dans la langue ne permet d'obtenir le sens du dérivé à partir de celui du radical qui a été isolé (cas O1). Une vignette n'est pas une petite vigne comme le laisserait penser l'application de la règle correspondant à R1. De même pour *baguette* ; quant à la signification de *cassette*, le patron correspondant à R9 prédit erronément 'casse de petite envergure'. Pour le deuxième type, O2, aucune segmentation n'est possible en synchronie et conséquemment aucune signification n'est dérivable par quelque patron actuel que ce soit. Certains des mots relevant de O1 et O2 peuvent néanmoins avoir été segmentables à des étapes antérieures de la langue et avoir été dérivés par des patrons ayant toujours cours. C'est le cas de *houlette* (< a. fr. *houler* 'lancer, jeter'), *galet* (< a. fr. *gal* 'caillou' < gaul.), *meurette* (< *muire* 'eau salée naturelle' < lat. *muria*), ou *clarinette* (< *clarin* 'hauboïs' cf. *clarine* 'cloche que portent les vaches'), etc. Ceci correspond au degré O3. Les mots restants sont ceux qui ne sont pas analysables en diachronie et qui n'ont jamais été dérivés. Relèvent du degré O4 *squelette* (< grec *skeletos*), les mots issus de noms propres (*silhouette*, *pamphlet*), les emprunts assimilés (*baguette*, *machette*, *mouffette*), les mots du fonds latin érodés et, pour ainsi dire, rechapés (*assiette*, *emplette*).

N°	PERIODE	ANALYSABLE	PATRON	EXEMPLE	TYPE
1	synchronie	oui	oui	<i>clochette, réformette, beurette</i>	T
2	synchronie	oui	non	a) <i>vignette</i> b) <i>galette</i> < a fr. <i>gal</i> c) <i>baguette</i> < it. <i>bacchetta</i>	O1
3	synchronie	non	non	a) <i>houlette, galet, clarinette</i> b) <i>squelette, silhouette, disette</i>	O2
4	diachronie	oui	oui	a) <i>vignette</i> b) <i>galette, meurette</i> c) <i>houlette, galet, clarinette</i>	O3
5	diachronie	non	non	a) <i>squelette</i> < grec b) <i>pamphlet, silhouette</i> c) <i>baguette</i> d) <i>disette, assiette</i>	O4

Tableau 3. Opacité / Transparence

Pour toutes les occurrences du corpus, nous avons distingué celles qui étaient transparentes (*clochette, réformette*), codées par T dans la base de données, celles qui étaient construites opaques (*houlette, clarinette*,

vignette) codées par CO et celles qui étaient non construites (*baguette, assiette, squelette*), codées par NC.

Le reste de l'article sera consacré à l'étude de la productivité comparée de *-et* et de *-ette* d'une part et des différents patrons d'autre part.

3. Méthodologie

La productivité d'un procédé morphologique ne peut être calculée ou estimée que pour un corpus donné. Par ailleurs les résultats obtenus dépendent fortement de ce corpus. Nous avons travaillé à partir d'un corpus d'archives électroniques du quotidien *Libération*. Ce corpus a une taille relativement importante puisqu'il couvre une période de cinq ans allant de janvier 1995 à décembre 1999. Il présente ainsi une bonne homogénéité du point de vue de la période et du genre, même s'il existe des variations entre les différentes rubriques du journal.

3.1 Extraction de la listes de noms en *-et* et *-ette*

Après avoir supprimé les fiches documentaires associées aux articles, le texte a été segmenté, étiqueté et lemmatisé à l'aide du système Cordial Analyseur (version 7) qui est l'un des meilleurs analyseurs actuellement disponibles et qui fonctionne aussi bien sous Windows que sous Linux. Le corpus segmenté contient 87 millions d'occurrences. Les résultats de l'analyseur comportent naturellement des erreurs, en particulier pour les mots absents de son dictionnaire. Ces dernières ne sont cependant pas pénalisantes car les mots inconnus sont étiquetés par défaut comme des noms communs et c'est justement ces derniers qui nous intéressent.

L'étiquetage du corpus est destiné à filtrer les formes verbales qui se terminent en *-ette* et *-ettes* comme *admette, permettes* ou *projettes* mais aussi les formes adjectivales comme *aigrelet, indiscret* ou *maigrelette*, et les noms propres comme *Juliette* ou *Rambouillet*. La lemmatisation permet, quant à elle, de supprimer les marques de pluriel. Notons que l'étude n'aurait pu être menée sans recours à des outils de traitement automatique car le corpus contient 617 000 unités qui comportent les finales *-et*⁹, *-ets*, *-ette* et *-ettes*.

Les occurrences des noms communs dont les lemmes se terminent par *-et* et par *-ette* ont été extraites du corpus. Deux listes ont ainsi été constituées, l'une pour les occurrences dont les lemmes se terminent en *-et* et l'autre pour celles dont les lemmes ont la finale *-ette*. Le nombre des occurrences de chacun de ces lexèmes a été calculé dans le corpus. Le tableau 4 présente les statistiques globales pour les deux listes :

	occurrences	lemmes	hapax
<i>-et</i>	213 265	570	242
<i>-ette</i>	57 176	575	155
total	270 441	1 145	397

Tableau 4. Statistiques pour les données brutes

⁹ Les occurrences de la conjonction *et* sont décomptées.

La différence entre les nombres d'occurrences des lemmes en *-et* et en *-ette* est due aux mots de fréquences élevées. Ainsi, les cinq types en *-et* les plus fréquents (*effet*, *projet*, *juillet*, *objet*, *budget*) totalisent 101 189 occurrences. Réciproquement, si on ne comptabilise que les types de fréquence inférieure à 1 000, les nombres d'occurrences de mots en *-et* et *-ette* deviennent plus comparables : 51 041 pour *-et* contre 34 387 pour *-ette*. Ce point est important car il détermine la validité de la comparaison de la productivité des deux suffixes (Gaeta & Ricca 2003) (cf. §4.1). Sans cela, il faudrait, comme le proposent ces auteurs, utiliser des sous-corpus de taille différente pour avoir un nombre d'occurrences proches pour chaque suffixe. Le tableau 4 montre également que les deux listes ont des nombres de lemmes et de hapax comparables.

3.2. Mesure standard de la productivité des affixes

L'une des hypothèses qui sous-tend ce travail est que *-et* et *-ette* sont des suffixes distincts, en particulier du point de vue de la forme, mais apparentés sémantiquement. (La nature et l'extension de cette parenté est laissée pour des études ultérieures.) Si c'est le cas, la productivité quantitative de ces suffixes devrait présenter des différences nettes. Pour les comparer, nous nous servons de la mesure proposée par (Baayen 1992) qui a été utilisée dans de nombreux travaux sur la productivité des procédés morphologiques (Baayen & Lieber 1991; Baayen & Renouf 1996; Plag, Dalton-Puffer & Baayen 1999). Cette mesure P est définie comme suit :

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

où n_i est le nombre d'hapax construits par un procédé morphologique (c'est-à-dire patron) donné et où N est le nombre total d'occurrences des lexèmes construits par de ce même procédé (pour une présentation plus détaillée, cf. Dal ici même). La mesure de Baayen est adaptée à notre étude car c'est d'abord un outil de comparaison de la productivité d'affixes concurrents ou proches. Nous l'utilisons pour comparer les suffixes *-et* et *-ette* globalement l'un à l'autre mais aussi de manière fine, en fonction de la catégorisation sémantique présentée en §§2.4-2.6.

3.3. Construction de la base de données

Les données brutes ne permettent de tirer aucune conclusion concernant la productivité de *-et* et de *-ette* : les critères de forme (sur la catégorie et sur la terminaison des lemmes) ne suffisent pas à identifier les mots construits au moyen des suffixes *-et* et *-ette*. Seul un dépouillement manuel des données permet de repérer les erreurs et les lemmes des lexèmes qui ne sont pas construits. Ce travail a donné lieu à la construction d'une base de données. L'analyse des occurrences a été effectuée en utilisant les contextes dans lesquels elles apparaissent. Nous avons aussi consulté des dictionnaires

comme le *Nouveau Petit Robert* et le *TLF* informatisé¹⁰ et des moteurs de recherche sur Internet (principalement Google^{TM11}).

Les lexèmes analysés peuvent être répartis en trois groupes : ceux qui ne vérifient pas les critères formels de catégorie (nom commun) et de forme (la finale doit être *-et* ou *-ette*) ; ceux qui ne sont pas construits ; ceux qui le sont (voir §2.8). Seuls les éléments du dernier groupe sont pris en compte pour calculer la productivité.

3.3.1. Éléments ne satisfaisant pas les critères formels

Les formes qui ne satisfont pas les critères formels sont des non-mots (fautes de frappe) ou des mots mal catégorisés par Cordial Analyseur.

1. Les fautes de frappe comme *bugdget*, *intêret*, *decret*, *ceette*, *médecineet.*, *instrumnet...* représentent une part non négligeable des entrées (86 lemmes en *-et* et 38 en *-ette*).
2. On trouve aussi, parmi les mots en *-et*, 36 acronymes relevant tous ou presque du domaine de l'informatique (*telnet...*) et d'Internet (*calvanet*, *globenet...*). Un seul acronyme en *-ette*, apparaît dans le corpus : *fadette* pour « facture détaillée ».
3. Les noms propres (*wallstreet*, *annette*, *babinet*, *pinochet*, *brossolette*, *villette...*) et certains noms de marque (*adminet*) écrits en minuscule constituent une autre catégorie d'éléments à éliminer (40 lemmes pour *-et* et 6 pour *-ette*). D'autres comme *garigette*, *mobyette* et *sanisette* ont en revanche été conservés.
4. Parmi les autres erreurs d'étiquetage, on trouve des formes verbales en *-ette* (comme *rétrojetette* ou *refeuilletette*), des adjectifs (*indiscret*, *simplet*, *sujette*, *simpletette...*), des éléments d'onomatopées comme *codet* ou *larirette*, des parties d'expression comme *lurette*, ou encore le déterminant *cette* étiqueté à tort comme nom commun.

Le tableau suivant présente les nombres d'occurrences et de lemmes qui ne satisfont pas les critères formels.

	occurrences	lemmes
<i>-et</i>	305	177
<i>-ette</i>	525	62
total	830	239

Tableau 5. Nombre de formes exclues par les critères formels.

On constate ainsi que si le nombre d'occurrences est négligeable (0,3%), ces entrées constituent néanmoins une partie importante des lemmes (21%).

3.3.2. Éléments ne satisfaisant pas le critère morphologique

Les lemmes qui satisfont aux critères formels ne sont cependant pas tous des lexèmes construits en *-et* ou en *-ette*. Parmi ces derniers se trouvent :

¹⁰URL : <http://zeus.inalf.fr/tlf.htm>

¹¹URL : <http://www.google.fr>

1. des mots en langues étrangères, essentiellement en anglais, qui apparaissent par exemple dans des titres d'œuvres cinématographiques, des extraits de chansons... (*feet, laundrette...*).
2. des emprunts à d'autres langues : anglais (*budget, junket, rocket*), italien (*morette, vedette*), latin savant (*effet, objet, pendet*), picard (*montinette*), provençal (*blanquette = « vin », farigoulette, nadalet*), catalan (*barrufet*), hébreux (*chaménet*), russe (*obrazet*). On trouve aussi dans le corpus des mots d'ancien français comme *soret*.
3. des mots construits à l'aide du suffixe homographe *-ette*, emprunté à l'italien, qui forme des noms désignant des ensembles de musiciens. Ce suffixe apparaît aussi dans des emprunts à l'anglais ayant la même signification : *tercet, quintette, sextet, sextette, septette, tentette...*
4. des mots construits à l'aide d'autres procédés : apocopes comme *amphet, perpet, perpette, cafet...* ; lexèmes préfixés ou composés comme *vidéocasette, eurovignette, supercalculette, antiprojet, superpréfet, téraoctet*¹² ; lexèmes construits par conversion comme *caquet* à partir de *caqueter* ou *projet* de *projeter* ; mots-valises comme *nuisette, midinette, netiquette...*

Ces lexèmes non construits par *-et* ou *-ette* représentent un nombre d'occurrences et de lemmes plus important encore que celui des candidats qui ne satisfont pas les critères formels. Le critère morphologique permet ainsi d'éliminer 60% des occurrences et 30% des lemmes.

	occurrences	lemmes
<i>-et</i>	152 493	233
<i>-ette</i>	21 112	116
total	173 605	349

Tableau 6. Nombre de formes exclues par les critères morphologiques

Ces formes sont conservées dans la base de données, d'une part parce qu'elles apparaissent dans d'autres corpus et qu'on pourra ainsi les repérer plus facilement ; d'autre part parce que quelques-unes des formes qui ne satisfont pas le critère morphologique présentent néanmoins les propriétés sémantiques des lexèmes construits en *-et* ou *-ette* (comme *alouette, gambette, maquette, tabouret*). Leur appartenance aux catégories décrites en §§2.4-2.6 est indiquée dans la base de données comme elle l'est pour les lexèmes-construits.

On observe ainsi que la majeure partie des données brutes (60% des occurrences et 51% des lemmes) décrites dans la base de données ne satisfont pas les critères formels ou morphologiques et doivent être décomptées pour les mesures de la productivité. Notre expérience rejoint en ce sens celle d'Evert et Lüdeling (Evert & Lüdeling 2001) qui montrent qu'il est impossible de mesurer la productivité des affixes de manière totalement automatique. Les outils de traitement automatique des langues apportent une aide précieuse pour l'exploration des corpus, mais la correction manuelle et l'analyse linguistique des résultats obtenus est impérative.

¹² Un milliard d'octets.

3.3.3. Lexèmes construits en *-et* et *-ette*

On peut distinguer parmi ces derniers deux sous-groupes selon que la nature de lexème construit est transparente (CT) ou opaque (CO ; cf. §2.8). Le tableau 7 indique le nombre d'occurrences et de lemmes des lexèmes construits.

	occurrences			lemmes		
	CT	CO	total	CT	CO	total
<i>-et</i>	8 594	29 141	37 735	110	42	151
<i>-ette</i>	23 387	6 730	30 117	340	56	393
total	31 981	35 871	67 852	450	98	546

Tableau 7. Nombre de lexèmes construits.

On constate que le nombre total d'occurrences pour *-et* et *-ette* est de même ordre de grandeur ; il est donc légitime de comparer la productivité de ces suffixes dans le corpus *Libération* (Gaeta & Ricca 2003).

Comparaison avec le TLF. Les hapax qui apparaissent dans un corpus donné, bien qu'étant de bons marqueurs de la productivité des affixes (c'est-à-dire de leur aptitude à former des mots nouveaux), ne sont pas nécessairement tous des productions lexicales nouvelles. Ce fait paraît lorsqu'on compare la liste des lemmes retenus avec les entrées du *TLF* (nomenclature et index). Le *TLF* comporte 596 noms masculins en *-et* et 754 noms féminins en *-ette*. Comme dans le cas des candidats extraits du corpus *Libération*, beaucoup de ces entrées ne sont pas des lexèmes construits transparents en *-et* ou *-ette*. Un bon nombre des lexèmes construits qui sont des hapax dans le corpus *Libération* sont en fait attestés dans le *TLF* (*castellet, coqueret, foutriquet, galoubet... gigolette, gloriette, gorgerette, gringalette...*). On ne peut donc pas les considérer comme des productions lexicales nouvelles.

	CT ∩ TLF	HL	HL ∩ TLF
<i>-et</i>	91	23	7
<i>-ette</i>	243	77	29
total	334	100	36

Tableau 8. Intersection entre le corpus *Libération* et les entrées du *TLF*.

Dans le tableau 8, 'CT' est l'ensemble des lemmes de lexèmes construits transparents dans le corpus *Libération*, 'HL' est l'ensemble des lexèmes construits transparents dans le corpus *Libération* qui sont des hapax et 'TLF' est l'ensemble des entrées correspondantes du *TLF*.

3.3.4. Format de la base de données

La base de données qui a été construite fournit pour chaque entrée les informations nécessaires pour le calcul de la productivité des affixes ainsi que des informations supplémentaires utiles pour d'autres utilisations de cette ressource. Les champs qui la composent sont :

1. la fréquence, c'est-à-dire le nombre d'occurrences du lexème dans le corpus ;
2. le lemme ;
3. le type, qui indique si le lexème est construit transparent (CT), s'il est construit opaque (CO ; voir §2.8) ou s'il ne vérifie pas les critères formels ou morphologiques. Dans ce cas, la cause du rejet est précisée ;
4. deux champs « interprétation » (Intrep-1, Interp-2) décrivent la ou les catégories sémantiques dont relève l'entrée. Ces catégories et la multi-appartenance qu'elles autorisent sont discutées en détail aux §§2.4-2.6 et §2.9 ;
5. le lexème-base est ensuite fourni pour les lexèmes CT et CO ;
6. un champ « commentaire », qui permet de décrire le sens du lexème ou de préciser certaines des informations nécessaires à son analyse ;
7. un champ « dictionnaire » indique dans quels dictionnaires le lexème est attesté (*Nouveau Petit Robert, TLF*).

Signalons que le nombre de lemmes est inférieur à celui du nombre des entrées de la base de données car une même forme lemmatisée peut subsumer des lexèmes différents (*caillette, canette, civette, couette, gigolette, mesurette, reinette, roussette*). Ces lemmes font de ce fait l'objet de plusieurs entrées. Par exemple *gigolette* a deux entrées :

1	gigolette	CT	R3		gigolo
1	gigolette	CT	R1	L5	gigot

4. Productivité

La productivité des suffixes *-et* et *-ette* a été calculée séparément pour les lexèmes CT et les lexèmes CO. Trois niveaux de catégorisation peuvent être distingués dans la base de données :

- la distinction entre lexèmes en *-et* vs. *-ette* est très sûre puisqu'elle est purement formelle ;
- la classification en lexèmes non construits, construits opaques et construits transparents est un peu moins sûre : elle est basée sur l'identification des lexèmes bases, sur l'intuition et sur les informations fournies par les dictionnaires ;
- la caractérisation relativement aux différentes catégories des pôles R et L nécessite une analyse plus fine et est davantage sujette aux variations intersubjectives.

De ce fait, les valeurs données pour les pôles R et L sont moins sûres que celles qui sont relatives à l'ensemble des lexèmes CT, CO et construits transparents ou opaques. Cette différence est accentuée par le fait que les différents lexèmes ont des occurrences en nombre très variables et quelques changements dans leur catégorisation peuvent induire des écarts importants

dans les valeurs de la mesure de productivité. Signalons par ailleurs que certains lexèmes CO n'ont pas été catégorisés davantage. De ce fait, le total pour les lexèmes CO est supérieur à la somme des sous-totaux R et L.

4.1 Lexèmes construits transparents

On observe pour les lexèmes construits transparents que le nombre d'occurrences pour *-et* est nettement inférieur à celui des occurrences pour *-ette*.

	hapax		occurrences		productivité	
	<i>-et</i>	<i>-ette</i>	<i>-et</i>	<i>-ette</i>	<i>-et</i>	<i>-ette</i>
R1	8	15	2 632	9 071	0,003 8	0,001 8
R2	1	4	74	2 802	0,011 6	0,001 4
R3	0	17	68	389	0	0,043 7
R4	1	0	976	1 039	0,001 0	0
R5	0	3	0	5	n.d.	0,6
R6	1	2	28	126	0,035 7	0,015 9
R7	2	0	13	51	0,153 8	0
R8	0	0	0	469	n.d.	0
R9	1	10	1	573	1	0,017 5
R10	2	1	5	84	0,4	0,011 9
total R	16	52	3 809	14 522	0,004 2	0,003 6
L1g	3	10	4 757	10 860	0,000 6	0,000 9
L1v	0	0	22	1 626	0	0
L2	0	3	0	175	n.d.	0,017 1
L3	2	6	708	544	0,002 8	0,011 0
L4	2	3	148	142	0,013 5	0,021 1
L5	6	11	74	71	0,081 1	0,154 9
total L	13	33	5 492	13 237	0,002 4	0,002 5
total	23	80	8 594	23 387	0,002 7	0,003 4

Tableau 9. Productivité des lexèmes CT.

Dans le tableau 9, la productivité est nulle pour les catégories qui n'ont pas d'hapax. C'est le cas de R3 et L1v pour *-et* de R4, R7, R8 et L1v pour *-ette*. La productivité n'est pas définie (notée 'n.d.') pour les catégories (R5, R8 et L2 pour *-et*) qui ne sont pas représentées dans le corpus. Par ailleurs, certaines valeurs de productivité sont très élevées du fait du nombre d'occurrences trop faible des lexèmes des catégories correspondantes : R9 (*attifet*) et R10 (*dramelet*, *scandalounet*, *scoopinet*) pour *-et* ; R5 pour *-ette* (*cuirette*, *jambonette*, *topazette*, *satinette*). Il s'agit là de biais introduits par le corpus et par les analyses décrites dans la base de données : ces valeurs ne sont donc pas significatives. Parmi les catégories dont les valeurs sont les plus fiables (à savoir celles qui ont un nombre suffisant d'occurrences et d'hapax), on voit que L5 est la plus productive (pour *-et* et pour *-ette*), suivie de R3 et de R9 (pour *-ette* seulement). Viennent ensuite L3 et L4 (pour *-et* et pour *-ette*), L2 et R10 (pour *-ette* seulement) que l'on peut

regrouper dans un second ensemble car le nombre de leurs hapax est plus faible.

Le tableau 9 indique que sur le corpus total, *-et* est plus productif que *-ette* pour les lexèmes relevant du pôle Référent et qu'il l'est moins pour ceux relevant du pôle Locuteur et pour l'ensemble des deux pôles. Cette différence provient cependant du fait que les lexèmes construits en *-ette* sont plus fréquents que ceux qui sont construits en *-et*. Il y a en effet 3,8 fois plus d'occurrences de lexèmes CT en *-ette* relevant du pôle R que d'occurrences de lexèmes CT en *-et*. Pour le pôle L, le rapport est de 1 pour 2,4. Or, la mesure de la productivité proposée par Baayen ne prend en compte que les occurrences des lexèmes construits au moyen du procédé morphologique auquel on s'intéresse. Dans le cas du corpus *Libération*, la productivité de *-et* pour le pôle R est donc mesurée relativement à un corpus d'occurrences 3,8 fois plus petit que celui sur lequel est mesurée la productivité de *-ette*. Chaque hapax de lexème en *-et* contribue ainsi 3,8 fois plus à la productivité de ce suffixe que ne le fait un hapax de lexème en *-ette* à la productivité de *-ette*.

Pour supprimer ce biais, (Gaeta & Ricca 2003) proposent d'utiliser pour les deux suffixes des sous-corpus qui contiennent les mêmes nombres d'occurrences pour les procédés à comparer. La productivité de *-et* pour le pôle R est calculée en considérant les 3 809 occurrences de lexèmes CT construits au moyen de ce suffixe et qui relèvent de ce pôle. En utilisant cette méthode, la productivité de *-ette* pour le pôle Référent doit être calculée relativement au même nombre d'occurrences. Nous avons donc projeté les analyses de la base de données sur les occurrences des lexèmes en *-ette*. La productivité de *-ette* peut ainsi être calculée en considérant uniquement les 3 809 premières occurrences de lexèmes CT en *-ette* du pôle Référent. Le rapport entre la productivité de *-et* et de *-ette* devient alors plus clair. Dans le corpus restreint, on trouve 33 hapax en *-ette*, ce qui donne une productivité de 0,008 7, supérieure à celle de *-et*. De même, il y a 25 hapax parmi les 5 492 premières occurrences de lexèmes CT en *-ette* appartenant au pôle Locuteur. La productivité de *-ette* passe ainsi à 0,004 6 pour les lexèmes de ce pôle. Pour l'ensemble des lexèmes CT en *-ette*, on trouve 55 hapax parmi les 8 594 premières occurrences ce qui donne une productivité de 0,006 4 dans ce corpus. Le tableau 10 reprend ces chiffres :

	hapax		occurrences	productivité	
	<i>-et</i>	<i>-ette</i>		<i>-et</i>	<i>-ette</i>
pôle R	16	33	3 809	0,004 2	0,008 7
pôle L	13	25	5 492	0,002 4	0,004 6
total CT	23	55	8 594	0,002 7	0,006 4

Tableau 10. Productivité avec des sous-corpus de tailles différentes.

Il apparaît ainsi clairement que, *-ette* est deux fois plus productif que *-et* pour les lexèmes CT tant pour le pôle Référent, pour le pôle Locuteur que pour l'ensemble des deux pôles.

4.2 Lexèmes construits opaques

Les lexèmes CO se caractérisent par le fait qu'ils appartiennent tous au français de longue date. On peut donc s'attendre à ce que, parmi ces lexèmes, les hapax soient absents. Cette hypothèse est globalement vérifiée, même si la taille nécessairement finie du corpus induit un biais, à savoir la présence de quelques hapax CO. Cet état de fait a deux conséquences qui apparaissent clairement dans le tableau 11 :

	hapax		occurrences		productivité	
	-et	-ette	-et	-ette	-et	-ette
R1	0	1	13 593	1 866	0	0,000 5
R2	1	0	1	148	1	0
R3	0	0	0	10	n.d.	0
R4	0	0	23	0	0	n.d.
R5	0	0	0	0	n.d.	n.d.
R6	0	0	0	0	n.d.	n.d.
R7	0	0	0	0	n.d.	n.d.
R8	0	0	0	0	n.d.	n.d.
R9	0	0	0	153	n.d.	0
R10	0	0	0	0	n.d.	n.d.
total R	1	1	13 617	2 177	0,000 1	0,000 5
L1g	1	1	7 926	1 655	0,000 1	0,000 6
L1v	0	0	152	73	0	0
L2	0	0	93	0	0	n.d.
L3	0	0	885	21	0	0
L4	0	0	0	314	n.d.	0
L5	0	0	23	725	0	0
total L	1	1	8 290	2 788	0,000 1	0,000 4
total	2	2	29 141	6 677	0,000 1	0,000 3

Tableau 11. Productivité pour les lexèmes construits opaques

D'une part, la productivité des suffixes *-et* et *-ette* pour les lexèmes CO est quasi-nulle : les deux dernières colonnes du tableau 11 contiennent des valeurs presque toutes nulles ou non définies. (Signalons que ces résultats sont « formels » car la mesure n'a de sens que relativement aux lexèmes CT.)

Ce tableau montre également qu'il existe un déséquilibre net entre les pôles Référent et Locuteur. R1 concentre la quasi-totalité des occurrences tant pour *-et* que pour *-ette* ; les autres catégories du pôle Référent ne contiennent des occurrences que pour l'un des deux suffixes. Cette complémentarité s'observe aussi pour le pôle Locuteur (L2 ; L3 ; L4 ; L5). Par ailleurs, L1(g) concentre la grande majorité des lexèmes CO aussi bien pour *-et* que pour *-ette*.

Comparaison avec les lexèmes CT. Les lexèmes de catégories R1 et L1(g) sont de loin les plus fréquents tant pour les lexèmes CT que CO. R1 et L1(g) constituent de ce fait des catégories « par défaut » pour *-et* et *-ette*. On observe d'autre part une certaine spécialisation de *-ette* (R2 ; R8 ; R9 ; R10) et dans une moindre mesure de *-et* (R4 ; L3). La comparaison des tableaux 11 et 13 permet également de voir que lexèmes CT et CO des catégories du

pôle Référent présentent des fonctionnements contrastés. La plupart ne sont pas représentées parmi les occurrences des lexèmes CO alors qu'elles le sont parmi celles des lexèmes CT. On peut donc formuler l'hypothèse qu'il y a eu pour ce pôle, en diachronie, une diversification de l'utilisation de *-et* et de *-ette*. On observe aussi que les lexèmes CO en *-et* sont cinq fois plus fréquents que ceux en *-ette* ; la situation est ainsi inverse pour les lexèmes CT. On peut donc émettre une deuxième hypothèse, à savoir que *-et* a été plus productif que *-ette* à certaines époques¹³.

5. Conclusion

Cet article constitue la première étude de la productivité des procédés de suffixation en *-et* et *-ette* menée sur un corpus journalistique au moyen des outils de mesure basés sur le nombre d'hapax mis au point par H. Baayen. Cette étude repose sur l'analyse détaillée du système de la dérivation en *-ET* en français proposée dans la deuxième partie. Cette analyse défend l'idée que ce système est organisé autour de deux grands pôles (Référent, Locuteur) et met en jeu une quinzaine de patrons de dérivation. Cette analyse était nécessaire pour pouvoir assigner chaque type morphologique décompté (chaque lemme) à un patron de dérivation, car aucun patron de dérivation unique ne peut fournir simultanément toutes les interprétations attestées¹⁴.

Les principaux résultats de cet article sont les suivants :

- La productivité de *-ette* est en général supérieure à celle de *-et*, ce qui justifie de ne pas considérer ces deux suffixes comme les exposants d'un même procédé morphologique. Ce point de vue est renforcé par le fait que les patrons de dérivation pour lesquels *-et* et *-ette* sont productifs ne sont pas identiques.
- Tous les patrons dérivationnels ne sont pas productifs au même degré. Certains sont improductifs (R8), d'autres très productifs (L5). Ce constat justifie le fait de ne pas traiter la suffixation en *-ET* comme un procédé unique. Corrélativement, la mesure de la productivité globale de cette suffixation n'a pas de sens, si ce n'est comme cumul des chiffres mesurant la productivité de chaque patron.

Ces résultats nous renseignent sur les zones actives du procédé et nous donnent une première image de son inscription dans la langue. Ils demeurent néanmoins très partiels et demandent à être étayés par d'autres travaux. Il faudra en particulier refaire les mesures sur d'autres corpus en faisant varier les caractères de taille, de période et genre pour neutraliser les effets liés au corpus particulier que nous avons utilisé et pour éviter d'avoir des mesures surévaluées à cause d'un nombre de types attestés trop faible. C'est seulement au terme de ces études qu'on pourra cerner les tendances

¹³ M. Plénat et M. Roché (Plénat & Roché 2001) ont eux aussi noté cette évolution pour laquelle ils proposent qu'elle ait été favorisée par un facteur morpho-phonologique, à savoir le fait que la consonne finale de *-et* a cessé d'être prononcée vers la fin du XVI^e siècle ; *-et* est alors devenu trop « léger » pour constituer un suffixe phonologiquement optimal.

¹⁴ Contrairement à ce qui se passe pour bon nombre de procédés de construction morphologique e.g. *-iser*, *-age1*, *-age2*, l'anglais *un-*, *-ness*. La situation de *-ET* ressemble, en plus complexe, à ce qu'on a avec la suffixation de *-ly* en anglais : le procédé forme des adjectifs ou des adverbes, mais seul le dernier est productif (Baayen & Renouf 1996).

lourdes du procédé de suffixation en *-ET*. Cependant, pour avoir une idée précise des caractéristiques de ce dernier, il sera nécessaire d'étudier comparativement la productivité d'autres procédés de construction morphologique en français.

6. Références

- Aliquot-Suengas Sophie. 1997. "Les suffixes **-ade** et **-é(e)**: une histoire de famille". In *Mot et grammaires*, Fradin B. & J.-M. Marandin (eds). 49-87. Paris: Didier Erudition.
- Baayen Harald. 1992. "Quantitative aspects of morphological productivity". *Yearbook of Morphology* 1991:109-149.
- Baayen Harald & Rochelle Lieber. 1991. "Productivity and English derivation: a corpus-based study". *Linguistics* 29 5:801-843.
- Baayen Harald R. & Antoinette Renouf. 1996. "Chronicling the **Times**: Productive lexical innovations in an English newspaper". *Language* 72:69-96.
- Bourdieu Pierre. 1979. *La distinction. Critique sociale du jugement*. Paris: Les Éditions de Minuit.
- Dal Georgette. 1997. *Grammaire du suffixe -et(te)*. Paris: Didier Erudition.
- Delhay Corinne. 1996. *Il était un "petit X". Pour une approche nouvelle de la catégorisation dite diminutive*. Paris: Larousse.
- Dressler Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and other Languages*. Berlin: Mouton de Gruyter.
- Evert Stefan & Anke Lüdeling. 2001. "Measuring morphological productivity: Is automatic preprocessing sufficient?" In *Proceedings of the Corpus Linguistics 2001 Conference*, Rayson P., A. Wilson, T. McEnery, A. Hardie & S. Khoja (eds). 167-175.
- Fradin Bernard. 1999. "La suffixation en **-ET** est-elle évaluative?" In *Morphologie des dérivés évaluatifs*, Corbin D., G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat & M. Roché (eds). 69-82. Villeneuve d'Ascq: UMR 8528 - CNRS & Université de Lille 3.
- Fradin Bernard. 2003. "Le traitement de la suffixation en **-et**". *Langages*.
- Gaeta Livio & Davide Ricca. 2003. "Italian prefixes and productivity: a quantitative approach". *Acta Linguistica Hungarica* 50:89-108.
- Grandi Nicola. 2002. *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*. Milano: Francoangeli.
- Jurafsky Daniel. 1996. "Universal tendencies in the semantics of the diminutive". *Language* 72:533-578.
- Kennedy Christopher & Louise McNally. 1999. "From Event Structure to Scale Structure: Degree Modification in Deverbal Adjectives". In *SALT IX*, Mathews T. & D. Strolovitch (eds). 163-180. Ithaca: CLC Publications.
- Lakoff George & Mark Johnson. 1980. *Metaphors We Live By*. Chicago / London: The University of Chicago Press.
- Plag Ingo, Christiane Dalton-Puffer & Harald Baayen. 1999. "Morphological Productivity across speech and writing". *English Language and Linguistics* 3:209-228.

- Plénat Marc & Michel Roché. 2001. Entre morphologie et phonologie: la suffixation décalée. Toulouse: Université Toulouse Le Mirail / ERSS.
- Stump Gregory T. 1991. "A paradigm-based theory of morphosyntactic mismatches". *Language* 67 4:675-725.
- van Marle Jaap. 1985. *On the Paradigmatic Dimension of Morphological Creativity*. Dordrecht: Foris Publications.

Annexe 1

Liste des hapax construits en *-ette* du corpus *Libération*. La deuxième colonne indique si le lexème est construit transparent (CT) ou construit opaque (CO). La troisième donne la ou les catégories sémantiques de lexème.

arielette	CT L1g	galonnette	CT R1	reniflette	CT R9
ariette	CT R2	giclette	CT R9	retroussette	CT R9
arthuisette	CT L2	gigolette	CT R1L5	riflette	CT L3'
bayonnette	CO L1g	gigolette	CT R3	ritonette	CT R3
bichounette	CT L5	gloriette	CO R1	roussette	CT L3
bimbette	CT R6	gorgerette	CT L1g	sablette	CT R1
bleuette	CT R2	gratounette	CT L1g	sanguette	CT L4
bobopoupette	CT L5	gravette	CT R1	saquette	CT L1g
bouillette	CT R2L1g	greluchette	CT L5	sifflette	CT R9
brindillette	CT R1	grévette	CT R9	soûlette	CT R9
caillette	CT L3	gringalette	CT R3	soupette	CT R9
caillette	CT R1	guignolette	CT R3	souricette	CT R3
carbichounette	CT L5	jambonnette	CT R5;L4	spationette	CT R3
carrelette	CT L1g	jeannette	CT R3	sudisette	CT R9
chatounette	CT L5	lancette	CT R1	suivette	CT R9
clonette	CT R2	machinette	CT R1	sumotorette	CT R3
cochette	CT R3	merlette	CT R3	tigette	CT R1
cochonnette	CT L5	mirette	CT L1g	topazette	CT R5
cosmonette	CT R3	montagnette	CT R1	tourniquette	CT L1g
crisette	CT R10	mouflette	CT R3	tranchette	CT R1
cuirette	CT R5	pépette	CT L5	usinette	CT R1
débarbouillette	CT L1g	périssolette	CT L2	vampirette	CT R3L5
dentelette	CT R1	pitchounette	CT L5	zizette	CT L5
diguette	CT R1	poivrette	CT L3		
dosette	CT R1	porcelette	CT R3		
doucette	CT L3	pulpette	CT R3		
fafette	CT R3	pulpinette	CT R3L5		
farfouillette	CT R9	quenouillette	CT L1g		
gallandette	CT L2	reinette	CT R6		

Annexe 2

Liste des hapax construits en *-et* du corpus *Libération*. La deuxième colonne indique si le lexème est construit transparent (CT) ou construit opaque (CO). La troisième donne la ou les catégories sémantiques de lexème.

attifet	CT L1g	dramelet	CT R10	portefeuillet	CT R1
banderolet	CT L1g;R1	flaconnet	CT R1	potelet	CT R1
biset	CT L3	flipounet	CT L5	poupounet	CT L5
bugounet	CT R2;L4	foutriquet	CT R9	scandalounet	CT R10;L5
cabasset	CO R2	gendrelet	CT R6	shortounet	CT R1;L5
clavinet	CT R1	îlet	CT R1	sitounet	CT L5
coqueret	CT L3	louchet	CO L1g	squalelet	CT R4
crémet	CT R7;L4	maset	CT R1		
curbelet	CT R7;L4	perret	CT L1g		

Résumé

Cet article esquisse tout d'abord une analyse du système de la suffixation en *-ET* en français. Cette analyse montre que ce système est organisé autour de deux grands pôles (Réfèrent, Locuteur) et qu'il met en jeu une quinzaine de patrons de dérivation. Sur cette base, l'article mène une étude de la productivité des procédés de suffixation en *-et* et *-ette* sur un corpus de cinq ans du journal *Libération* au moyen des outils de mesure mis au point par H. Baayen. Deux résultats partiels émergent : la productivité de *-ette* est supérieure à celle de *-et*, ce qui justifie de ne pas considérer ces deux suffixes comme les exposants d'un même procédé morphologique ; tous les patrons dérivationnels ne sont pas productifs au même degré, ce qui implique que la productivité globale de ce procédé ne peut être autre chose que la somme de la productivité obtenue pour chaque patron.

Abstract

This article first sketches an analysis of the system of *-ET* suffixation in French. This analysis shows that the *-ET* suffixation involves approximately fifteen patterns of lexeme formation organized around two poles (the Referent and the Speaker's pole). Building on this analysis, the article examines the productivity of the derivational suffixes *-et* and *-ette* in the French newspaper *Libération* in a five years period. Formal tools provided by H. Baayen are used to gauge the degree of productivity of both affixes. Two partial results are worth mentioning : *-ette*'s degree of productivity exceeds *-et*'s one ; this result supports the idea that these suffixes must not be considered as exponents of the one and same morphological process. Not all derivational patterns are productive to the same degree, which means that it is meaningless to speak of the global degree of productivity of *-ET* suffixation, except as the sum of the degrees of productivity obtained for each pattern.