

Règles et paradigmes en morphologie informatique lexématique

Nabil Hathout(1) Fiammetta Namer(2)

(1) UMR CLLE-ERSS, Toulouse

(2) UMR ATILF-Université Nancy2, Nancy

nabil.hathout@univ-tlse2.fr, fiammetta.namer@univ-nancy2.fr

Résumé.

Les familles de mots produites par deux analyseurs morphologiques, DériF (basé sur des règles) et Morphonette (basé sur l'analogie), appliqués à un même corpus lexical, sont comparées. Cette comparaison conduit à l'examen de trois sous-ensembles :

- un sous-ensemble commun aux deux systèmes dont la taille montre que, malgré leurs différences, les approches expérimentées par chaque système sont valides et décrivent en partie la même réalité morphologique.
- un sous-ensemble propre à DériF et un autre à Morphonette. Ces ensembles (a) nous renseignent sur les caractéristiques propres à chaque système, et notamment sur ce que l'autre ne peut pas produire, (b) ils mettent en évidence les erreurs d'un système, en ce qu'elles n'apparaissent pas dans l'autre, (c) ils font apparaître certaines limites de la description, notamment celles qui sont liées aux objets et aux notions théoriques comme les familles morphologiques, les bases, l'existence de RCL « transversales » entre les lexèmes qui n'ont pas de relation d'ascendance ou de descendance.

Abstract.

The word families produced by two morphological analyzers of French, DériF (rule-based) and Morphonette (analogy-based), applied on the same lexical corpus have been compared. The comparison led us to examine three classes of relations:

- one subset of relations that are shared by both systems. It shows that, despite their differences, the approaches implemented in these systems are valid and describe, to some extent, one and the same morphological reality.
- one subset of relations specific to DériF and another one to Morphonette. These sets (a) give us informations on the characteristics proper to each system, and especially on what the other system is unable to produce; (b) they highlight the errors of one system, in so that they are absent from the results of the other; (c) they reveal some of the limits of the description, especially the ones related to theoretical objects and concepts such as morphological family, base or the existence of transverse LCR (lexeme construction rules) between lexemes that are not ascendant nor descendant of each other.

Mots-clés : morphologie constructionnelle ; analyse automatique ; règles ; analogie ; familles morphologiques ; comparaison ; synergie

Keywords: Word formation ; automatic analysis ; rules ; analogy ; morphological families ; comparison ; synergy.

1 Introduction

Cet article présente une comparaison entre deux ressources morphologiques très différentes : l'analyseur DériF et le réseau morphologique Morphonette. DériF (Namer, 2009) est un analyseur qui implémente des règles de construction de lexèmes (RCL) établies et mises au point manuellement. Leur application est contrôlée par un ensemble d'exceptions qui permettent de prendre en compte efficacement l'ensemble des irrégularités qui se sont accumulées au cours de

l'évolution du lexique. Les analyses produites par DériF indiquent pour chaque mot construit l'opération qui a servi à le construire et sa base ou ses éléments de composition. Morphonette est pour sa part basé sur un modèle relationnel paradigmatique du lexique (Hathout, 2009, 2011) dans lequel les mots sont identifiés par leur famille et leur série morphologiques. Ce réseau est construit de façon totalement automatique en utilisant une mesure de la proximité morphologique, l'analogie et trois critères qui permettent d'éliminer les paradigmes et les relations les moins sûres.

Les deux systèmes sont comparés sur le même corpus : le lexique flexionnel TLF_{nome}¹. Les analyses morphologiques produites par les deux systèmes ne sont pas de même nature et ne peuvent être comparés directement. Notamment, Morphonette ne fournit pas de relation dérivationnelle entre les lexèmes, mais seulement les paradigmes qui les contiennent. En revanche, les analyses de DériF sont plus fines et permettent de reconstruire les familles et les séries morphologiques des lexèmes. Ces paradigmes constituent l'un des plans de comparaison possible pour les deux systèmes. Cette comparaison fait clairement apparaître leurs différences tant d'un point de vue qualitatif que quantitatif (cf §2 vs §3). Dans ce premier travail, nous avons limité la comparaison à un seul type de paradigmes : les familles morphologiques (§4). Les séries morphologiques seront traitées ultérieurement. Les familles sont décrites de façon explicite dans Morphonette. En revanche, il faut les reconstruire à partir des résultats de DériF en réalisant une fermeture transitive des relations de dérivation et de composition en prenant garde à ne pas connecter les composés qui ne partagent aucun élément de composition (§4.1). La comparaison des familles ainsi extraites des deux systèmes conduit à l'obtention de trois sous-ensembles, décrits quantitativement (§4.2) et qualitativement (§4.3) : il s'agit d'un sous-ensemble commun, d'un sous-ensemble propre à DériF et d'un autre propre à Morphonette.

Les objectifs de ce travail sont multiples : identifier les convergences qui existent entre les deux systèmes ; analyser les résultats pour identifier les compromis qui ont été réalisés dans chaque système et les hypothèses et généralisations sur lesquels ils s'appuient ; mettre en évidence certaines améliorations dont les systèmes pourrait profiter mutuellement l'un de l'autre ; viser une convergence plus forte ; faire ressortir des questions linguistiques qui n'émergent que grâce à la comparaison des systèmes.

2 DériF

DériF s'applique à des mots considérés hors-contexte (non fléchis). Il met en jeu des règles établies suivant des critères linguistiques : l'analyse proposée permet de réunir les ascendants du mot, et la relation qu'entretient chaque mot avec son ou ses ascendants immédiats : sa base dans le cas d'une conversion ou d'une affixation, ses composants dans le cas d'une composition (1).

L'analyse est donc tendanciellement orientée. En cas de dérivation récursive (Spencer, 1988 ; Becker, 1993) comme en (2), le système décide de la solution la plus motivée linguistiquement : la base *désenneiger* est préférée à *enneigement*. Si plusieurs analyses sont possibles (du fait de l'existence de plusieurs interprétations) toutes les solutions sont fournies (cf 3a vs 3b). La logique du système de règles a pour vocation de prédire le sens morphologiquement construit des néologismes. Pour cela, les hypothèses formulées vont de la situation la plus particulière au cas par défaut : celui-ci est censé refléter la production néologique, illustrée sous (4) par l'analyse d'un nom inexistant mais possible.

Étant donné la lexicalisation, l'opacification d'un certain nombre d'éléments du lexique attesté (stockés de longue date), l'application des règles est soumise à l'examen du contenu de listes d'exceptions. Par exemple, l'inscription de *pension* parmi les exceptions évite au système de l'analyser comme déverbal de *penser* (5a), sur le modèle de *pression*, dérivé de *presser* (5b).

- (1) enneigement/NOM : (enneigement/NOM, enneiger/VERBE, neige/NOM)
- (2) désenneigement/NOM : (désenneigement/NOM, désenneiger/VERBE, enneiger/VERBE, neige/NOM)
- (3) réenneigement/NOM :
 - a. (réenneigement/NOM, enneigement/NOM, enneiger/VERBE, neige/NOM)
 - b. (réenneigement/NOM, réenneiger/VERBE, enneiger/VERBE, neige/NOM)
- (4) a. reschtroumpfement/NOM : « schtroumpfement réitéré »
 b. reschtroumpfement/NOM : « (Action - résultat de l'action) de reschtroumpfer »
- (5) a. pression/NOM : (pression/NOM, presser/VERBE)
 b. pension/NOM : (pension/NOM)

Les informations qui annotent les mots analysés bénéficient des caractéristiques de DériF : étant donné que le principe fondamental est le fruit de l'information linguistique, il est possible de produire (i) la paraphrase par laquelle se définit le mot par rapport à sa base (6) et (ii) les traits traduisant les contraintes imposées par chaque RCL (7). Les autres

¹ Ce lexique a été construit à partir de la nomenclature du TLF. Il correspond à la version 1.0 de Morphalou disponible sur le site du CNRTL (www.cnrtl.fr/lexiques/morphalou/).

résultats (p. ex. la liste des ascendants successifs du mot analysé, reliés chacun à une RCL, comme l'exemple sous (1)) sont représentés sous la forme de listes.

(6) chanteur/NOM : « (Agent habituel - Auteur exceptionnel - Instrument) de chanter »

(7) a. chanter/VERBE: [aspect = dynamique, sous_cat = <NPagent, ...>]

b. chanteur/NOM: [concret = oui, comptable = oui]

3 Morphonette

Morphonette est un réseau lexical du français basé sur une conception relationnelle et paradigmatique de la morphologie. Dans ce lexique, les propriétés morphologiques sont décrites par la position des lexèmes dans le réseau, positions identifiées par les paradigmes qui les contiennent. Par exemple, la position d'un dérivé comme *modifiable* peut être minimalement décrite par sa famille morphologique qui rassemble les lexèmes *modifier*, *modification*, *modificateur*, *modificatif*, *modifiant*, *modifieur*, *immodifiable*, etc. et par sa série qui contient l'ensemble des dérivés en *-able* : *agaçable*, *agitabile*, *chevauchable*, *définissable*, *différenciable*, *rechargeable*, *réconciliable*, *soutenable*, etc.

Dans sa version actuelle, Morphonette est construit à partir de TLFnome (Hathout, 2010), qui est également le lexique de test et de référence de DériF. Il est composé de **filaments**, c'est-à-dire de triplets $(m, p, s_p(m))$ où m est une entrée, p , un membre de la famille morphologique de m et $s_p(m)$ la sous-série morphologique de m relativement à p . $s_p(m)$ est l'ensemble des mots du lexique qui se trouvent dans une relation similaire à celle que m entretient avec p . En d'autres termes, $u \in s_p(m)$ s'il existe v tel que $m:p=u:v$ (i.e. tels que m, p, u, v forment une analogie). Par exemple, l'un des filaments de *modifiable* dans Morphonette est :

(8) (*modifiable*, *modificateur*, {*amplifiable*, *glorifiable*, *identifiable*, *justifiable*, *clarifiable*, *mystifiable*, *rectifiable*, *sanctifiable*, *simplifiable*, *spécifiable*, *unifiable*, *vérifiable*})

Dans ce lexique chaque entrée a autant de filaments que de membres dans sa famille. Si certains filaments se recouvrent en grande partie et devront à terme être fusionnés, d'autres décrivent des propriétés différentes d'un même mot. Par exemple, *gazouillard* a dans sa famille le nom *gazouillarde* (9a) et le verbe *gazouiller* (9b). Sa sous-série relativement au verbe est présentée en (9b). Elle contient des dérivés comme *bafouillard* ou *vasouillard* qui peuvent être rattachés aux verbes *bafouiller* ou *vasouiller*. En revanche sa sous-série relative au nom féminin *gazouillarde* (9a) contient en plus les noms pour lesquels il existe un féminin en /aRd/ comme *gaillard*, *montagnard*, *trouillard* ou *savoyard* mais pas de verbe correspondant. On a ainsi pour ce mot des sous-séries distinctes, l'une correspondant à la propriété d'être un nom déverbal et l'autre à celle d'être associée à un féminin en /aRd/.

(9) a. (*gazouillard*, *gazouillarde*, {*babillard*, *becquillard*, *béquillard*, *braillard*, *douillard*, *égrillard*, *fripouillard*, *gaillard*, *grenouillard*, *grognard*, *justiciard*, *montagnard*, *prétentiard*, *savoyard*, *citrouillard*, *trouillard*, *vadrouillard*, *vasouillard*})

b. (*gazouillard*, *gazouiller*, {*bafouillard*, *douillard*, *grenouillard*, *citrouillard*, *vadrouillard*, *vasouillard*, *ventrouillard*})

La construction de Morphonette vise ainsi à découvrir certaines des relations morphologiques entre les mots du lexique de départ. Leur identification est réalisée par étapes en réduisant successivement l'espace de recherche des membres de la famille et de la série de chaque mot. Ces étapes sont :

(a) la sélection pour chaque entrée d'un voisinage constitué des 100 mots qui sont morphologiquement les plus proches de l'entrée. Ces voisinages sont définis par une mesure qui rapproche les mots qui partagent le plus grand nombre possible de propriétés formelles (n -grammes de lettres ou de phonèmes) les plus spécifiques possible (Hathout, 2009). La mesure est calculée au moyen d'un algorithme de propagation d'activation inspiré des travaux de Gaume (2004). Cette mesure est différente de la distance d'édition de Levenshtein.

(b) la découverte de l'ensemble des analogies formelles qui s'établissent entre les entrées et les mots de leurs voisinages (Lepage, 1998; Stroppa et Yvon, 2006). Afin de réduire la complexité de cette exploration, ces analogies sont identifiées sur la base de signatures d'édition (séquence d'opérations d'édition) calculées pour chaque couple (m, v) où m est une entrée et v un de ses voisins. Les couples qui partagent les mêmes signatures forment des analogies (Hathout, 2009).

(c) l'utilisation de trois critères pour séparer les familles des séries et pour sélectionner les paradigmes qui comportent le moins d'erreurs possible. Le premier est un critère catégoriel qui permet de séparer les relations familiales des relations sérielles pour les analogies dont les 4 éléments ne sont pas de la même catégorie. Le deuxième est un critère portant sur la taille des sous-séries : deux mots qui appartiennent à la même famille morphologique participent normalement à un nombre élevé d'analogies. On élimine par conséquent les filaments dont la sous-série contient moins de 10 mots. Le troisième critère porte sur le degré d'agrégation des sous-séries (Watts et Strogatz, 1998) : les éléments qui se trouvent dans l'une des sous-séries d'un mot donné ont tendance à apparaître ensemble dans une des autres sous-séries du réseau. On calcule donc pour chaque élément de chaque sous-série un taux d'agrégation (rapport entre le nombre de triangles et

le nombre de triples qui contiennent cet élément et les autres membres de la sous-série) et on élimine ceux dont le niveau d'agrégation est inférieur à 0,66.

Les critères utilisés, ajoutés aux restrictions imposées lors des deux premières étapes, permettent de disposer d'un réseau très fiable mais, en contre partie, relativement petit.

4 Comparaison

4.1 Formatage des résultats

La comparaison de DériF et de Morphonette est réalisée dans un premier temps sur les familles morphologiques. Nous envisageons de l'étendre dans un second temps aux séries. Dans Morphonette, la famille d'un lexème peut être directement récupérée de la liste de ses filaments. Dans le cas de DériF, seules les relations entre un mot et ses ascendants sont indiquées. Nous avons donc reconstruit des familles en ajoutant des relations transversales entre les descendants de chaque mot. Pour les affixations et les conversions, ces relations sont obtenues en calculant leurs fermetures symétriques et transitives. En revanche, pour les compositions, les relations transversales doivent être limitées aux seuls descendants de chaque lexème. A l'heure actuelle, les chiffres présentés ci-dessous n'incluent pas les descendants de composants qui ne sont pas des lexèmes, et notamment des formants néoclassiques (par exemple, la relation qu'entretient dans DériF *hépatique/A* et *hépat/N*=foie* n'est pas comptabilisée ici). 9 582 entrées sont concernées par cette restriction.

4.2 Résultats quantitatifs

La comparaison des deux systèmes porte sur les 96 456 entrées de TLFnome de catégories nom, verbe, adjectif et adverbe. Le tableau de gauche ci-dessous donne le nombre d'interprétations calculées par DériF pour les entrées de TLFnome. Celui de droite fournit quelques statistiques sur Morphonette. Les différences entre ces tableaux reflètent bien celles qui existent entre les structures des résultats que produisent les deux systèmes.

DériF					Morphonette		
Nb interprétations	0	1	plusieurs	total	Entrées	Relations familiales	Relations sérielles
Effectif	60 882	34 722	853	96 456	29 310	96 107	1 160 098

Nous avons extrait pour chaque entrée analysée par DériF l'ensemble des relations qu'elle a avec ses ascendants (par exemple *harmonie* et *harmoniser* pour *harmonisation*). Ces relations sont ensuite complétées par des relations transversales entre tous les descendants de l'entrée. On obtient ainsi après symétrisation 269 710 couples de mots qui appartiennent à la même famille. Ces couples ont été comparés aux 96 107 relations familiales qui se trouvent dans Morphonette. Le tableau suivant présente cette comparaison des relations familiales d'un point de vue quantitatif (D représente l'ensemble de celles dans DériF et M de celles de Morphonette). La taille de l'intersection montre que, malgré leurs différences importantes, les approches expérimentées par chaque système sont valides et décrivent en partie la même réalité morphologique.

Relations familiales	D	M	$D \cap M$	$D \setminus M$	$M \setminus D$
Effectif	269 710	96 107	26 061	189 849	70 046

4.3 Comparaison qualitative

L'examen du contenu des trois sous-ensembles nous donne un certain nombre de renseignements sur la qualité des analyses de chaque système. En particulier, les sous-ensembles propres à chaque analyseur dénoncent autant ce qui manque à l'autre que ce que lui-même doit améliorer. Tout d'abord, le sous-ensemble propre à Morphonette permet de dégager trois grands types de résultats, exposés ci-dessous.

DériF ne retrouve pas les ascendants indirects (*prédateur/N*, *prédation/N*) s'ils n'ont pas d'ancêtre commun : *°prédateur/V* n'est pas attesté (sur la morphologie paradigmatique, voir p. ex. Booij, 1997, 2008). DériF est aussi totalement muet en ce qui concerne certaines RCL, ce dont Morphonette sert de révélateur (p. ex. les noms en *-euse* ou les adjectifs en *-ant*, cf. Anscombre (2000))

Morphonette réunit des unités de composition néoclassique qui partagent le composant déterminé (*zoophage/N*, *anthropophage/N*) ou déterminant (*zoographe/N*, *zoophage/N*), alors que DériF ne relie les composés que lorsqu'ils mettent en jeu le même composant déterminé (*zoographe/N*, *biographe/N*), ou deux composants déterminés entretenant une relation lexicale spécifique : p. ex. synonymie (*zoovore/N*, *zoophage/N*) ; sur la composition dite savante ou néoclassique en français, cf. (Dal, 2008), sur son analyse automatique, cf. (Namer, 2007).

Morphonette produit des résultats instanciant des cas d'analogie formelle sémantiquement non valides. Les familles directes calculées par DériF sont l'un des moyens de détecter ces fausses analogies. Prenons par exemple le couple (*brasserie/N*, *brassière/N*). Il a été forgé sur le même modèle que (*poissonnerie/N*, *poissonnière/N*). Or l'équivalence formelle de (*brasserie/N*, *brassière/N*) avec (*poissonnerie/N*, *poissonnière/N*) est invalidée par l'absence d'analogie sémantique entre ces deux couples, absence qui peut-être identifiée automatiquement par l'examen des familles d'ascendants directs de ces quatre noms, calculée par DériF. En effet, dans la paire (*poissonnerie/N*, *poissonnière/N*) les familles en jeu pour chaque élément sont (*poissonnerie/N*, *poissonnier/N*, *poisson/N*) et (*poissonnière/N*, *poissonnier/N*, *poisson/N*). Ces deux familles constituent une structure de la forme (X_{erie_N} , X_{ier_N} , X_N) et ($X_{ière_N}$, X_{ier_N} , X_N) où l'ancêtre commun X est le nom *poisson*. Cette même structure se retrouve avec p. ex. (*chocolaterie/N*, *chocolatière/N*), (*fauconnerie/N*, *fauconnière/N*), (*laiterie/N*, *laitière/N*). Ces couples sont donc en relation d'analogie formelle et d'identité structurelle familiale. En revanche, les familles directes associées à *brasserie/N* et *brassière/N* sont (*brasserie/N*, *brasser/V*, *bras/N*) et (*brassière/N*, *bras/N*). Leur structure est doublement différente : la première famille fait intervenir un verbe (*brasser/V*) et la seconde ne comporte que deux membres : (*brassière/N*, *bras/N*).

Les couples qui n'apparaissent que dans DériF peuvent être répartis en plusieurs groupes. Le premier contient 14 572 couples qui se trouvent dans une relation de conversion nom-adjectif ou adjectif-nom : les filaments dont l'entrée et le membre de la famille ont la même forme ont été en effet supprimés de Morphonette. Le second groupe est celui des mots préfixés. Ces derniers sont éliminés de Morphonette par les trois critères (cf. §3). La comparaison met ainsi en évidence l'une des lacunes principales de ce réseau. Les raisons de l'absence de préfixation dans les filaments de Morphonette découlent : (1) de la tendance de ces relations à être isocatégorielles, (2) du fait du nombre plus faible de dérivés préfixés que de dérivés suffixés et (3) du fait de la forte dispersion des relations transversales que les préfixés entretiennent avec les membres plus distants de leurs familles. Sont également absentes de Morphonette les allomorphies comme dans les couples (*abrégéable/A*, *abréger/V*) qui se trouvent éliminés lors de la deuxième étape, du fait du nombre réduit de couples qui comportent cette variation. Signalons également qu'un petit nombre de couples, qui ne sont présents que dans DériF, sont des erreurs comme (*léguer/V*, *légion/N*) ou (*lecteur/N*, *lion/N*) qui ne sont pas morphologiquement apparentés en synchronie.

5 Conclusion

Nous avons présenté un ensemble de résultats encore exploratoires issus d'une comparaison de deux systèmes d'analyse morphologique, appliqués au même ensemble de données. Nous avons vu que la présence d'un élément, aussi bien que son absence, dans un sous-ensemble propre à un système renseigne sur son propre bruit, autant que sur le silence de l'autre système. Également, la comparaison permet de rapprocher des heuristiques apparemment très différentes : notamment, les listes d'exception de DériF constituent la formulation linguistique à laquelle correspond le taux d'agrégation de Morphonette. Outre les conclusions prévisibles concernant le positionnement de chaque système par rapport à l'autre, suivant différents critères habituellement utilisés en TAL (robustesse, bruit, silence, quantité d'information produite, etc.), cette confrontation aboutit à deux types de résultats remarquables, en terme de « retours » sur les hypothèses théoriques fondatrices des deux systèmes :

- D'un point de vue épistémologique, l'existence d'une intersection de résultats communs à DériF et à Morphonette couvrant 10% des analyses, démontre que deux théories aussi à l'opposée l'une de l'autre que sont le courant morphologique à base de règles orientées, et l'approche paradigmatique, fondée sur l'analogie permettent d'aboutir pour une partie de leurs résultats à des conclusions convergentes. En outre, les résultats complémentaires à cette intersection sont tout aussi précieux que l'est le contenu de celle-ci. En effet, une fois notre système perfectionné, nous pouvons raisonnablement nous attendre à ce que ces résultats propre à l'un ou l'autre système constituent une démonstration empirique des limites de l'autre approche : ni la méthode à base de règles, ni celle fondée sur l'analogie ne peut se passer de l'autre.
- D'un point de vue de la validité même des hypothèses théoriques, les résultats erronés fournis par Morphonette et absents de DériF révèlent des critères permettant d'affiner les principes du raisonnement analogique qui associe une unité x_1 à une unité x_2 sur le modèle formel observable entre y_1 et y_2 : pour vérifier que la relation sémantique entre x_1 et x_2 est elle aussi analogue à celle qui relie y_1 à y_2 , on se sert de la structure des familles directes de x_1 , x_2 , y_1 et y_2 . À l'analogie des formes, les résultats obtenus automatiquement permettent d'associer **l'analogie des familles**, pour aboutir finalement à des projections en terme d'analogie sémantique. En retour, de ces nouvelles associations analogiques on peut dériver de nouvelles règles morphologiques, : à l'image du schéma $Xeur_N / Xion_N$ illustré au

§.4.2, ces **nouvelles règles** vont alors alimenter le système constructionnel existant sous forme de relations le plus souvent non orientées et/ou non binaires.

Finalement, ces premiers résultats laissent à penser que les améliorations futures de DériF et Morphonette vont conduire à un accroissement de leur intersection et donc à une convergence de plus en plus importante, au fur et à mesure que chaque système aura effectué les ajustements nécessaires dictés par l'autre. A terme, nous serons en mesure d'offrir aux utilisateurs un lexique qui cumule les qualités de chacun des deux systèmes, et donc des deux raisonnements théoriques sous-jacents, tout en évitant leurs inconvénients respectifs.

Ce lexique va être conçu sous la forme d'un réseau et dont les arcs seront valués par les relations sémantiques (à partir des pseudo-définitions des mots analysés engendrées par DériF, cf. §.2). En ce qui concerne les relations indirectes valides, p. ex. (*poissonnerie/N, poissonnière/N*) ou (*prédateur/N, prédation/N*), elles seront intégrées au réseau, en fonction de leur fréquence, avec leur sens construit indirectement, de manière à compléter le réseau par des arcs tissés et annotés entre des éléments d'une même famille lexicale, qui ne sont pas en relation d'ascendance directe.

Références

- Anscombe, Jean-Claude. 2000. Un problème de sémantique lexicale: l'interprétation active/passive des adjectifs verbaux participes en position d'épithète. In *Le passif - Etudes Romanes 45*, ed. Lene Schøester, 237-260. Copenhagen: Université de Copenhagen.
- Becker, Thomas. 1993. Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. *Yearbook of Morphology 1992*:1-27.
- Booij, Geert. 1997. Autonomous morphology and paradigmatic relations *Yearbook of Morphology 1996*:35-53.
- Booij, Geert. 2008. Paradigmatic Morphology. In *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, ed. Bernard Fradin, 29-38. Amsterdam / Philadelphia: John Benjamins.
- Dal, Georgette. 2008. L'analogie dans le domaine du lexique construit : un retour ? *1er Congrès Mondial de Linguistique Française*, Paris:1575-1587.
- Gaume, Bruno. 2004. Balades aléatoires dans les Petits Mondes Lexicaux. *I3. Information - Interaction - Intelligence*, 4(2).
- Hathout, Nabil. 2009. Acquisition of morphological families and derivational series from a machine readable dictionary. In Fabio Montermini, Gilles Boyé et Jesse Tseng, éd., *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, Somerville, Mass. : Cascadilla Proceedings Project.
- Hathout, Nabil. 2010. *Morphonette: a morphological network of French*. Préprint. <http://hal.archives-ouvertes.fr/hal-00485503/PDF/morphonette-en.pdf>
- Hathout, Nabil. 2011. Une approche topologique de la construction des mots : propositions théoriques et application à la préfixation en *anti-*, In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon et Marc Plénat, *Des unités morphologiques au lexique*, pp. 251–318. Hermes Science-Lavoisier : Paris.
- Lepage, Yves. 1998. Solving analogies on words: An algorithm. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and of the 17th International Conference on Computational Linguistics*, volume 2, pp. 728–735, Montréal.
- Namer, Fiammetta. 2007. Morphosémantique pour l'appariement de termes dans le vocabulaire médical : approche multilingue. *Traitement Automatique des Langues 46*:157-181.
- Namer, Fiammetta. 2009. *Morphologie, Lexique et TAL : l'analyseur DériF*: TIC et Sciences cognitives. London: Hermes Sciences Publishing.
- Spencer, Andrew. 1988. Bracketing paradox and the English Lexicon. *Language 64*:663-682.
- Stroppa, Nicolas et François Yvon. 2006. Du quatrième de proportion comme principe inductif : une proposition et son application à l'apprentissage de la morphologie. *Traitement automatique des langues*, 47(2):33-59.
- Watts, Duncan J. et Steven H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442.