

## La base lexicale Démonette : entre sémantique constructionnelle et morphologie dérivationnelle

Nabil Hathout<sup>1</sup>, Fiammetta Namer<sup>2</sup>

(1) UMR 5263 CLLE/ERSS, CNRS & Université Toulouse Le Mirail, Toulouse

(2) UMR 7118 ATILF, CNRS & Université de Lorraine, Nancy

Nabil.Hathout@univ-tlse2.fr, Fiammetta.Namer@univ-lorraine.fr

**Résumé.** Démonette est une base de données lexicale pour le français dont les sommets (entrées lexicales) et les arcs (relations morphologiques entre les sommets) sont annotés au moyen d'informations morpho-sémantiques. Elle résulte d'une conception originale intégrant deux approches radicalement opposées : Morphonette, une ressource basée sur les analogies dérivationnelles, et DériF, un analyseur à base de règles linguistiques. Pour autant, Démonette n'est pas la simple fusion de deux ressources pré-existantes : cette base possède une architecture compatible avec l'approche lexématique de la morphologie ; son contenu peut être étendu au moyen de données issues de sources diverses. L'article présente le modèle Démonette et le contenu de sa version actuelle : 31 204 verbes, noms d'action, noms d'agent, et adjectifs de propriété dont les liens morphologiques donnent à voir des définitions bi-orientées entre ascendants et entre lexèmes en relation indirecte. Nous proposons enfin une évaluation de Démonette qui comparée à VerbaCTION obtient un score de 84% en rappel et de 90% en précision.

**Abstract.** Démonette is a lexical database whose vertices (lexical entries) and edges (morphological relations between the vertices) are annotated with morpho-semantic information. It results from an original design incorporating two radically different approaches: Morphonette, a resource based on derivational analogies and DériF, an analyzer based on linguistic rules. However, Démonette is not a simple merger of two pre-existing resources: its architecture is fully compatible with the lexematic approach to morphology; its contents can be extended using data from various other sources. The article presents the Démonette model and the content of its current version, including 31,204 verbs, action nouns, agent nouns and property adjectives, where morphological links between both direct ascendants and indirectly related words have bi-oriented definitions. Finally, Démonette is assessed with respect to VerbaCTION with a recall of 84% and a precision of 90%.

**Mots-clés :** Réseau lexical. Morphologie dérivationnelle. Famille morphologique. Sémantique lexicale. Français

**Keywords:** Lexical Network. Derivational morphology. Morphological family. Lexical semantics. French.

### 1 Introduction

Cet article présente le réseau morpho-sémantique Démonette. Cette ressource comporte plus de 30 000 noms, verbes et adjectifs, dont les connections réalisent les relations morphologiques directes ou indirectes entre mots de la même famille dérivationnelle<sup>1</sup>. Les éléments de ce réseau sont des lexèmes typés sémantiquement ; les arcs sont valués par les relations morpho-sémantiques qui s'établissent entre ces mots ; ces relations sont regroupées en fonctions de leurs définitions abstraites. Ce réseau est le fruit d'un travail qui s'inscrit dans la continuité d'une étude préliminaire présentée dans (Hathout, Namer, 2011) qui montrait la complémentarité de deux ressources morphologiques dérivationnelles conçues selon des principes théoriques opposés, DériF et Morphonette, et comment elles peuvent être combinées.

La suite de l'article est organisée comme suit. §2 présente le modèle de Démonette, et notamment la conception cumulative du sens lexical construit (§2.2), le typage de ces mots construits, ainsi que les gloses<sup>2</sup> et leur abstraction

<sup>1</sup> Une famille dérivationnelle est un ensemble de mots apparentés directement ou indirectement. Elle comporte généralement un mot simple (morphologiquement indécomposable) et plusieurs mots construits. Les mots construits se répartissent entre composés (POISSON-CHAT ← POISSON, CHAT) et dérivés par suffixation (CHANTEUR ← CHANTER), par préfixation (IMPUR ← PUR) ou par conversion (SCIER ← SCIE). Dans cet article, il ne sera question que de dérivation.

<sup>2</sup> Une glose est la paraphrase en langue naturelle du mot construit par rapport au parent morphologique (i.e. au membre de la famille dérivationnelle) auquel il est relié.

(§2.3). Le §3 détaille la construction proprement dite du réseau lexical. Le §3.1 décrit le corpus de travail. Les §3.2 et 3.3 présentent respectivement DériF et Morphonette ainsi que la manière dont nous en avons extrait les informations destinées à alimenter Démonette. Une caractérisation qualitative et quantitative de ce lexique est proposée au §3.4. Suit en §4 une évaluation partielle de Démonette relativement au lexique Verbaction<sup>3</sup>. Nous abordons ensuite en §5 différentes questions liées à l'extension de cette nouvelle ressource et à la possibilité de découvrir et de caractériser des relations dérivationnelles qui composent Démonette sans passer par DériF ni Morphonette.

## 2 Le modèle Démonette

### 2.1 Une nouvelle ressource : justification

Un très grand nombre d'outils d'analyse morphologique, tant flexionnelle que constructionnelle ont été développés au cours des deux dernières décennies. Les travaux récents les plus visibles ont plutôt été réalisés par les informaticiens pour répondre à des besoins de décomposition morphologique et pallier les limites des lexiques. Leur objectif est le plus souvent le découpage des mots en morphèmes. Certains systèmes opèrent en plus un étiquetage des morphèmes permettant de regrouper leurs variantes. Ces systèmes utilisent tous des méthodes d'apprentissage automatique. Ils sont relativement indépendants des langues, mais ont tendance à produire de meilleurs résultats pour les langues à morphologie concaténative comme l'anglais, le français ou l'allemand. Citons notamment *Linguistica* (Goldsmith, 2001), *Morphessor* (Creutz, Lagus, 2005) ou l'analyseur de Bernhard (2006). Plusieurs de ces travaux ont été développés autour des campagnes Morpho-Challenge dont l'un des enseignements est que, contrairement à d'autres tâches comme l'annotation morpho-syntaxique, les performances des analyseurs morphologiques statistiques est globalement trop faible pour améliorer significativement les applications dans lesquelles ils seraient intégrés. On trouve à côté de ces systèmes statistiques des analyseurs plus « linguistiques » comme DériF (Namer, 2009) ; pour un panorama général, voir par exemple (Bernhard et al., 2011 ; Namer, 2013).

Parallèlement, on observe qu'aucun lexique dérivationnel de taille suffisante n'a été développé dans les 20 dernières années et que, depuis la création et la diffusion de CELEX (Baayen et al., 1995), aucune ressource similaire n'a été réalisée ou mise à disposition des linguistes, psycholinguistes et linguistes informaticiens. Ce manque est surtout crucial pour les langues romanes, notamment le français, l'italien et l'espagnol. Le constat était déjà fait il y a plus de 15 ans par Dal et al. (1999). Il reste inchangé.

Certaines ressources apportent certes des réponses partielles à ce besoin. Le réseau lexical *JeuxDeMots* (Lafourcade, Joubert, 2013) contient des relations morphologiques regroupées en classes dont certaines sont typées sémantiquement au moyen d'étiquettes comme 'verbe-action', 'action-verbe', etc. Cette ressource, constituée par les participants à plusieurs jeux en ligne bénéficie d'une couverture lexicale très large mais présente un défaut de systématisme dans les relations lexicales codées et notamment dans les relations morphologiques. Des relations morpho-sémantiques sont également présentes dans les réseaux WordNet de plusieurs langues comme l'anglais (Fellbaum, Miller, 2003), le tchèque (Pala, Hlaváčková, 2007) ou le turc (Bilgin et al., 2004). Dans le Princeton WordNet 3.0, par exemple, les relations morphologiques ne sont pas sémantiquement typées, et ne connectent que certains synsets (dont tous les mots ne sont pas concernés). On trouve par ailleurs au niveau des synsets des gloses du sens construit morphologiquement, mais qui ne mettent pas formellement en jeu le sens du synset de la base. Ces relations sont en revanche présentes dans le réseau lexical *FrameNet* (Ruppenhofer et al., 2002) qui comporte des gloses « abstraites » similaires à celle de Démonette (cf *infra*). Comme nous allons le voir, ces deux ressources partagent par ailleurs une organisation du sens lexical en « constellation » centrée autour des prédicats verbaux.

À terme, l'apport de Démonette sera triple. Il permettra d'une part d'offrir des **analyses morphologiques fiables** pour une langue qui ne dispose toujours pas d'un véritable lexique dérivationnel : la fiabilité est l'un des points forts de DériF (cf §3.1) dont Démonette bénéficie directement. La deuxième contribution de Démonette sera la construction d'une **ressource lexicale dérivationnelle** de grande taille<sup>4</sup>. La troisième est une structure lexicale en réseau qui coïncide parfaitement avec l'**approche lexématique** de la morphologie lexicale moderne (pour le français, cf entre autres Fradin, 2003 ; Roché et al., 2011). L'architecture de Démonette illustre en effet comment sémantique et morphologie constructionnelle peuvent s'articuler dans ce type d'approche.

### 2.2 Réseau morphologique dérivationnel

Démonette présente une architecture relativement originale. C'est un réseau lexical implémenté sous la forme d'un graphe où chaque mot est connecté à une partie des membres de sa famille dérivationnelle. Les mots construits sont

<sup>3</sup> <http://redac.univ-tlse2.fr/lexiques/verbaction.html>

<sup>4</sup> Dans sa version actuelle, Démonette inclut environ 15% de la nomenclature d'un dictionnaire comme le *Trésor de la Langue Française* (TLF).

reliés à leur base (par des **relations « directes »**) et à un sous-ensemble des autres mots construits de leur famille dérivationnelle (par des **relations « indirectes »**). Cette organisation reprend en partie celle de Morphonette (§3.2). Par exemple, les arcs en (1) relient les membres la famille de *essorer*. Les relations « directes » comme *essorage – essorer* sont notée en gras et les relations « indirecte » comme *essorage – essoreuse* en maigre.

- (1) **essorage – essorer** ; essorage – essoreur ; essorage – essoreuse ;  
**essoreur – essorer** ; essoreur – essorage ; essoreur – essoreuse ;  
**essoreuse – essorer** ; etc.

Les sommets qui représentent les mots contiennent des informations sur la catégorie, les traits morphosyntaxiques et éventuellement, le type d'affixation ayant permis de les construire. Par exemple, le sommet qui représente *essorage* indique qu'il s'agit d'un nom masculin construit par une suffixation en *-age*. Démonette offre ainsi une représentation morphologique complète des mots qu'elle décrit. Il fournit en outre un typage sémantique de ces mots et une description de leur **sens morphologique**.

## 2.3 Sémantique cumulative

L'objectif de Démonette est à terme de fournir pour chaque couple de mots du TLF, reliés directement ou indirectement, une annotation bi-orientée, de manière à ce que chaque mot construit soit défini relativement à chacun des mots avec lesquels il entretient une parenté morphologique. Cet objectif est plus facilement atteignable quand la relation est directe, c'est-à-dire entre un dérivé et sa base, grâce notamment aux gloses du sens du dérivé fournies par DériF (cf §3.1). En revanche le traitement des relations indirectes réclame davantage de réflexion (voir §5). Si on considère par exemple la famille partielle donnée en (2), l'étiquetage sémantique de l'arc entre *activateur* et *activation* est justifié (d'autant qu'il est reconstitué spontanément par les locuteurs, qui voient en *activateur* le nom désignant l'auteur d'une *activation*, et, réciproquement, dans le nom *activation*, l'acte réalisé par *l'activateur*) alors qu'il semble moins légitime entre le nom *activateur* et l'adverbe *activement*.

- (2) (actif, activer, activateur, activatrice, activation, activiste, activisme, activement, ...)

### 2.3.1 Des descriptions sémantiques redondantes

Chaque relation morphologique apporte une contribution au sens des mots qu'elle connecte. Ces sens élémentaires se cumulent alors pour produire le sens global. On peut considérer que le sens global d'une relation entre deux parents, e.g. *momifiable* et *momie* (3c) se déduit de la composition des relations élémentaires que chaque parent entretient avec son ascendant direct, ici *momifiable* et *momifier* d'une part (3a) et *momifier* et *momie* de l'autre (3b) :

- (3) a. *momifiable*/ADJ ← *momifier*/VERBE : « que l'on peut momifier »  
 b. *momifier*/VERBE ← *momie*/NOM : « transformer en momie »  
 c. *momifiable*/ADJ ← *momie*/NOM : « que l'on peut transformer en momie »

De même, on peut généralement calculer, l'un relativement à l'autre, le sens de deux mots complexes qui ont la même base : *momifiable* peut être défini relativement à *momifier* (3a) et *momifier* est connecté à *momification* par une relation que l'on peut gloser comme en (4a). En conséquence, *momifiable* peut tirer son sens de *momification* (4b), et réciproquement (4c).

- (4) a. *momification*/NOM ← *momifier*/VERBE « action de momifier »  
 b. *momifiable*/ADJ ← *momification*/NOM : « à qui il est possible d'appliquer la momification »  
 c. *momification*/NOM ← *momifiable*/ADJ : « acte applicable à ce(lui) qui est momifiable »

Les sens élémentaires des mots complexes apparentés sont ainsi interconnectés par le partage de prédicats et d'arguments. C'est cette faculté qui est exploitée dans Démonette : chaque mot y reçoit autant de définitions qu'il a de connections avec ses parents morphologiques. Dans sa version présentée ici, les descriptions sémantiques de Démonette sont exprimées en langage semi-naturel dans le but de rendre plus explicites ces interconnexions.

### 2.3.2 Séries dérivationnelles et typage sémantique des unités lexicales

Les mots et les relations sémantiques qui les connectent peuvent être regroupés en classes qui correspondent à des types sémantiques. Cette classification repose sur un niveau d'abstraction qui met en jeu des variables typées illustrées dans la table 1 et qui nous permet de comparer entre elles les unités et leur sens. Dans les représentations abstraites, le symbole @ correspond au prédicat verbal à partir duquel se définissent les noms et adjectifs qui lui sont apparentés ; de même, on note @FCT le nom ou l'adjectif qui remplit la fonction sémantique FCT vis-à-vis du prédicat @.

Variable typée	Signification	Séries dérivationnelles	Exemples
@ACT	Nom d'activité	Xment, Xage, Xion	<i>lavage, gonflement, exclusion</i>
@AGM	Nom d'agent masculin	Xeur	<i>danseur, sauveteur</i>
@AGF	Nom d'agent féminin	Xeuse, Xrice	<i>danseuse, sculptrice</i>
@PROP	Adjectif de propriété	Xif	<i>combatif</i>
@	Prédicat	–	<i>combattre, laver, danser, sculpter, gonfler, exclure</i>

TABLE 1 : typage sémantique des unités lexicales de Démonette

Par exemple, le dérivé *danseur* dénote l'(AG)ent (M)asculin du prédicat @ réalisé par le verbe d'action *danser*. Son étiquette est @AGM. Plus généralement, le type sémantique d'un mot *Y* appartenant à une famille dérivationnelle constituée autour d'un prédicat *X* est déterminée par la **série dérivationnelle** de *Y*, c'est-à-dire l'ensemble des mots construits par la même règle de dérivation que *Y*. Cette série dérivationnelle fournit ainsi l'identifiant du rôle sémantique que *Y* exerce vis-à-vis de *X*.

Actuellement, Démonette contient les membres de 7 séries dérivationnelles et leurs bases. Nous y trouvons une partie des verbes qui dénotent des actions, les séries dérivationnelles déverbiales des noms d'**action** en *-age*, *-ment* et *-ion* (dénotant les mêmes actions que leur verbe de base) et des noms d'**agent** en *-eur*, *-euse* et *-rice*, (réfèrent à des humains susceptibles de réaliser ces actions) et les adjectifs de **propriété** en *-if* (liés aux potentialités d'actualisation du prédicat verbal). Tous ces modes de formation sont facilement identifiables, productifs, et sémantiquement réguliers. Ils sont traités par les deux systèmes qui alimentent Démonette. De plus, les parentés indirectes y sont nombreuses et induisent des connections transverses intéressantes, cf (5), faciles à gloser, comme en (6) pour la relation entre *ravitailleur* et *ravitaillement*.

- (5) a. ravitailler, ravitaillement, ravitailleur, ravitailleuse  
b. capter, captage, captation, captatif, captateur, capteur  
c. décanter, décantage, décantation, décantement, décanteur
- (6) *ravitailleur*: (Agent masculin habituel - Auteur masculin exceptionnel - Instrument) de l'activité liée au *ravitaillement*

Dans certaines familles comme en (7), le verbe est absent. Les radicaux savants /predat/ et /odit/, communs aux membres de ces familles respectives sont la trace laissée par un prédicat en latin qui n'a pas d'héritier verbal en français contemporain. Néanmoins, les familles (7a) et (7b) sont composées de membres « légitimes » des séries dérivationnelles que nous avons sélectionnées, étant donné les propriétés formelles de ces mots et les caractéristiques sémantiques qui leur sont corrélées (*audition* dénote une action, *prédateur*, une classe d'agents, *auditif*, une propriété, etc.).

- (7) a. prédateur, prédation, prédatrice  
b. audition, auditif, auditeur, auditrice

### 2.3.3 Annotations sémantiques des relations entre unités lexicales

Les types sémantiques présentés ci-dessus sont utilisés comme paramètres dans les définitions associées aux relations entre les dérivés et leurs bases (relations directes, comme *fonder* → *fondateur*) ou entre des couples de mots partageant un même ascendant (relations indirectes, comme *fondateur* → *fondation*). Les **définitions abstraites** associées aux relations directes et indirectes sont établies selon les trois principes suivants :

1. Chaque membre d'une relation est représenté par son **type sémantique**.
2. Dans une relation directe entre un dérivé et sa base, le dérivé est défini par rapport au sens de sa base (définition dite **orientée**). Réciproquement, on indique par une définition **orientée inverse** la paraphrase du sens de la base relativement à celui du dérivé.
3. Une relation indirecte connecte deux mots complexes. Chacun y est défini par rapport à l'autre, donnant lieu à deux définitions **transversales**, symétriques l'une de l'autre, qui exploitent la possibilité de combiner les sens élémentaires de ces deux mots.

Ces principes permettent d'annoter uniformément —i.e. au moyen de deux définitions symétriques— l'ensemble des relations qui existent entre les couples de mots morphologiquement apparentées dans le graphe. Les définitions orientées, extraites des gloses générées par DériF (cf §3.1), servent à l'élaboration des autres annotations, orientées inverses et transversales. L'application des trois principes est illustrée dans les tables 2 et 3. La formulation de la

définition en colonne 2 fait systématiquement intervenir le prédicat @ ; le symbole ‘OU’ rend compte de la polysémie de ces mots construits. Notons que les relations indirectes peuvent s’établir entre des mots qui portent la même étiquette sémantique comme les dérivés concurrents (*lavage / lavement ; ausculteur / auscultateur*) ou les correspondants féminins / masculins du même agent (*danseuse / danseur*).

Schéma de relation directe dérivé/base	Définition orientée dérivé ← base	Définition orientée inverse base ← dérivé	Exemple
Xeur <sub>N</sub> /X <sub>V</sub>	(Agent masculin habituel OU Auteur masculin exceptionnel OU Instrument) de @	Réaliser l'activité dont l'agent masculin est le @AGM	marcheur / marcher
Xeuse <sub>N</sub> /X <sub>V</sub> ; Xrice <sub>N</sub> /X <sub>V</sub>	(Agent féminin habituel OU Auteur féminin exceptionnel OU Instrument) de @	Réaliser l'activité dont l'agent féminin est le @AGF	sculptrice / sculpter
Xage <sub>N</sub> /X <sub>V</sub> ; Xment <sub>N</sub> /X <sub>V</sub> ; Xion <sub>N</sub> /X <sub>V</sub>	(Action OU Résultat de l'action) de @	Réaliser le/la @ACT	abaissement / abaisser
Xif <sub>A</sub> /X <sub>V</sub>	En rapport avec l'acte de @	Manifester le fait d'être @PROP	combatif / combattre

TABLE 2 : Définitions orientées implémentées dans Démonette

Schéma de relation indirecte dérivé <sub>1</sub> /dérivé <sub>2</sub>	Définition transversale de dérivé <sub>1</sub> ← dérivé <sub>2</sub>	Définition transversale de dérivé <sub>2</sub> ← dérivé <sub>1</sub>	Exemple
(Xion <sub>N</sub> ; Xage <sub>N</sub> ; Xment <sub>N</sub> )/Xeur <sub>N</sub>	action pratiquée par @AGM	agent masculin ou instrument du @ACT	distracted / distracteur
(Xion <sub>N</sub> ; Xage <sub>N</sub> ; Xment <sub>N</sub> )/ (Xeuse <sub>N</sub> ; Xrice <sub>N</sub> )	action pratiquée par @AGF	agent féminin ou instrument du @ACT	fondation / fondatrice
Xif <sub>A</sub> /(Xage <sub>N</sub> ; Xment <sub>N</sub> ; Xion <sub>N</sub> )	qui permet la @ACT	action de ce qui est @PROP	déterminatif / détermination
Xif <sub>A</sub> /Xeur <sub>N</sub>	qui caractérise l'activité pratiquée par @AGM	celui dont l'activité est @PROP	administratif / administrateur
Xif <sub>A</sub> /(Xeuse <sub>N</sub> ; Xrice <sub>N</sub> )	qui caractérise l'activité pratiquée par @AGF	celle dont l'activité est @PROP	spoliatif / spoliatrice
(Xeuse <sub>N</sub> ; Xrice <sub>N</sub> )/Xeur <sub>N</sub>	celle qui a pour correspondant masculin le @AGM	celui qui a pour correspondant féminin la @AGF	danseuse / danseur
Xage/(Xion ; Xment) ; Xment/Xion	@ACT	@ACT	ruminement / rumination
Xeur <sub>N</sub> /X'eur <sub>N</sub>	@AGM	@AGM	activateur / activateur

TABLE 3 : Liste des définitions transversales implémentées dans Démonette

### 3 Données, méthodes et résultats

La structure générale de Démonette permet d’organiser d’une manière originale les informations morphologiques présentes dans les analyses de DériF et dans Morphonette. Pour permettre leur alignement et faciliter leur intégration, DériF a été appliqué sur le corpus à partir duquel Morphonette a été construit : TLFnome<sup>5</sup>. Le graphe Démonette est donc construit à partir d’un ensemble de sommets unique et connu à l’avance.

<sup>5</sup> L’ATILF distribue Morphalou, une version XML de TLFnome : [www.cnrtl.fr/lexiques/morphalou/](http://www.cnrtl.fr/lexiques/morphalou/). TLFnome est un lexique créé à partir de la nomenclature du dictionnaire TLF. Il comporte 97 000 lemmes et se caractérise par une qualité quasi parfaite grâce aux très nombreuses révisions manuelles dont il a bénéficié.

### 3.1 DériF

DériF (Namer, 2009, 2013) est un analyseur qui implémente des règles de construction morphologiques établies et mises au point manuellement, par des linguistes. Il se distingue par deux caractéristiques principales :

1. Il est fiable grâce notamment à un ensemble d'exceptions permettant de prendre en compte efficacement les irrégularités qui se sont accumulées au cours de l'évolution du lexique.
2. Il propose pour chaque dérivation une glose du sens construit, similaire à celles de WordNet (cf §2.1).

DériF analyse des lemmes munis d'une catégorie grammaticale. Il met en jeu des règles établies suivant des critères linguistiques : l'analyse proposée (ou les analyses, quand plusieurs sont possibles) réunit les ascendants du mot, et la relation constructionnelle de ce dernier avec son ou ses ascendants immédiats, sous la forme d'une paraphrase, cf §2.3.1. Pour les affixations et les conversions, l'ascendant immédiat est la base comme en (8). De par sa conception, DériF est aussi capable d'analyser des néologismes. Il prédit leur base et calcule leur sens construit en fonction de leur structure morphologique et leur catégorie grammaticale (en (9), *schtroumpfement* est un nom suffixé en *-ment*).

- (8) enterrement/NOM : (enterrement/NOM, enterrer/VERBE, terre/NOM) :  
"(Action OU résultat de l'action) d'enterrer"
- (9) schtroumpfement/NOM : (schtroumpfement/NOM, schtroumpfer/VERBE)  
"(Action OU résultat de l'action) de schtroumpfer"

DériF se distingue des autres analyseurs morphologiques par sa capacité à prendre en compte la lexicalisation et notamment l'opacification de certains mots du lexique enregistré. Ce phénomène est géré au moyen de listes d'exceptions qui bloquent l'application des règles pour ces mots. Par exemple, l'inscription de *pension* parmi les exceptions évite au système de l'analyser comme déverbal de *penser* (10a), sur le modèle de *pression*, dérivé de *presser* (10b).

- (10) a. pension/NOM : (pension/NOM)  
b. pression/NOM : (pression/NOM, presser/VERBE)

Les analyses de DériF servent à la fois à créer la structure du graphe de Démonette —la liste des sommets et des arcs— et à en calculer les attributs. Les analyses obtenues en appliquant DériF à TLFnome permettent de construire un premier graphe dans lequel les sommets sont les entrées de TLFnome et les arcs sont les relations dérivationnelles directes entre les mots construits et leurs ascendants directs (bases ou éléments de composition). Les sommets du graphe qui forment le corpus de Démonette sont ensuite alors en 3 temps : 1/ sélection des entrées de noms construits par l'une des 6 suffixations déverbales ; 2/ ajout des entrées des verbes de base correspondants ; 3/ ajout des entrées des adjectifs en *-if* dont la base est un nom ou un verbe préalablement sélectionné. Dans un second temps, les familles dérivationnelles sont munies de relations « indirectes ». La structure de graphe ainsi obtenue répond alors aux spécifications présentées en §2.1 et peut être directement intégrée à Démonette.

Outre la structure du graphe, les analyses de DériF permettent d'annoter les relations directes de Démonette au moyen des paraphrases (ou gloses) qui définissent les dérivés par rapport à leur base. D'autres attributs sont également déduits des analyses de DériF :

1. Les types sémantiques des sommets qui représentent des mots construits sont calculés à partir de la catégorie grammaticale (pour les verbes) ou de l'affixation utilisée pour construire le dérivé (pour les noms et adjectifs). Une unité lexicale peut théoriquement appartenir à plusieurs classes sémantico-aspectuelles correspondant à ses différentes acceptions. Dans les prochaines versions de Démonette intégrant davantage d'entrées, les mots pourront avoir plusieurs types associés à différentes relations dérivationnelles.
2. La définition abstraite de la relation directe est calculée à partir de la paraphrase concrète correspondante.
3. Les définitions abstraites transversales sont calculées en fonction des types sémantiques des sommets connectés.
4. À partir de ces définitions, on détermine les instances concrètes des relations indirectes.

### 3.2 Morphonette

Morphonette est un réseau lexical du français basé sur une conception relationnelle et paradigmatique de la morphologie (Hathout, 2011). Dans ce lexique, les propriétés morphologiques d'un mot sont décrites par les paradigmes qui le contiennent. Par exemple, les propriétés d'un dérivé comme *modifiable* peuvent être minimalement décrites par la famille dérivationnelle qui rassemble les mots *modifier*, *modification*, *modificateur*, *modificatif*, *modifiant*, *modifieur*, *immodifiable*, etc. et par la série des dérivés en *-able* : *agaçable*, *agitable*, *chevauchable*, *définissable*, *différenciable*, *rechargeable*, *réconciliable*, *soutenable*, etc.

Morphonette a été construit à partir de TLFnome. Il est composé de **filaments**, c'est-à-dire de triplets  $(m, p, s_p(m))$  où  $m$  est une entrée,  $p$  est un membre de la famille dérivationnelle de  $m$  et  $s_p(m)$  est la sous-série dérivationnelle de  $m$

relativement à  $p$ .  $s_p(m)$  est l'ensemble des mots qui se trouvent dans une relation similaire à celle que  $m$  entretient avec  $p$ . En d'autres termes, un mot  $u$  appartient à  $s_p(m)$  s'il existe un mot  $v$  tel que  $m : p = u : v$  (i.e. tel que  $m, p, u, v$  forment une analogie). L'exemple (15) présente le filament de l'adjectif *modifiable* pour  $p = \textit{modificateur}$ .

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

Dans Morphonette, une entrée a autant de filaments qu'il y a de membres dans sa famille dérivationnelle. Certains de ces filaments se recouvrent en grande partie. D'autres filaments décrivent des propriétés différentes d'un même mot. Par exemple, *travailleur* a dans sa famille le nom *travailleuse* et le verbe *travailler*. Sa sous-série relativement au verbe contient des dérivés comme *ravageur* ou *cisailleur* qui peuvent être rattachés aux verbes *ravager* ou *cisailler*. En revanche sa sous-série relative au nom féminin *travailleuse* contient en plus les noms comme *deuilleur* ( $\leftarrow$  *deuil*) ou *volaille* ( $\leftarrow$  *volaille*) pour lesquels il existe un féminin en /øz/ (*deuilleuse*, *volailleuse*) mais pas de verbe correspondant. *Travailleur* appartient donc à deux sous-séries différentes, l'une correspondant à la propriété d'être un nom déverbal et l'autre à celle d'être associée à un féminin en /øz/.

L'un des intérêts des filaments est qu'ils décrivent à la fois les relations directes entre un dérivé et sa base, indirectes avec les autres membres de sa famille et ses relations avec les membres de ses sous-séries. Les relations entre mots de la même famille sont utilisées pour alimenter Démonette. Morphonette ne donnant aucune information explicite sur les opérations qui ont permis de construire les dérivés, celles-ci sont recalculées en reconstituant les couples de mots qui partagent le même radical, comme par exemple en (16), puis en réunissant ceux qui instancient la même relation de parenté morphologique, comme en (17) entre les noms en *-euse* et les adjectifs en *-if*.

- (16) ('bouilleur/NOM', 'bouillage/NOM')
- (17) collectif/ADJ:collectionneuse/NOM = dissertatif/ADJ:disserteuse/NOM =  
nocif/ADJ:noceuse/N = perfectif/ADJ:perfectionneuse/NOM = portatif/ADJ:porteuse/NOM =  
possessif/ADJ:possesseuse/NOM = sélectif/ADJ:sélectionneuse/NOM

### 3.3 Contenu de Démonette

Dans sa version actuelle, Démonette comporte 31 204 relations décrivant chacune un couple de mots morphologiquement apparentés. Ces couples appartiennent aux séries dérivationnelles présentées au §2.2.2, fournies soit par DériF (21 556 couples issus des familles dérivationnelles regroupant l'ensemble des ascendants des mots construits sur un prédicat verbal), soit par Morphonette (9 648 relations, toutes indirectes).

MOT1	MOT2	Or	SEM1	SEM2	REL DIR1	REL INDIR1	REL DIR2	REL INDIR2
administratif /ADJ	administratrice /NOM	D	@PROP	@AGF	En rapport avec l'acte de @	qui caractérise l'activité pratiquée par @AGF	Agent féminin ou instrument de @	celle dont l'activité est @PROP
administrateur /NOM	administrer /VERBE	D	@AGM	@	(Agent masculin habituel - Auteur masculin exceptionnel - Instrument) de @		Réaliser l'activité dont l'agent féminin est le @AGM	
admonitrice /NOM	admonition /NOM	M	@AGF	@ACT		agent féminin ou instrument du @ACT		Action pratiquée par @AGF

TABLE 4 : Échantillon de Démonette

Une entrée de Démonette comporte 9 champs (table 4) : MOT1 et MOT2 sont les membres du couple décrit ; ORIGINE renseigne sur la provenance de l'entrée (Morphonette ou DériF) ; SEM1 et SEM2 enregistrent respectivement le type sémantique de MOT1 et MOT2, RELATION INDIRECTE1 est la définition transversale de MOT1 par rapport au sens de MOT2 ; RELATION INDIRECTE2 est celle de MOT2 relativement au sens de MOT1. Enfin, si MOT1 et MOT2 entretiennent une relation indirecte (ligne 1), RELATION DIRECTE 1 et 2 définissent MOT1 et MOT2 par rapport à leur ascendant verbal commun, alors que, si MOT1 est un dérivé de MOT2 ou

vice-versa (ligne 2), RELATION DIRECTE 1 et 2 enregistrent les définitions orientée et orientée inverse appropriées. Dans le second cas, les champs RELATION INDIRECTE 1 et 2 ne sont pas renseignés. 8 318 relations de Démonette proviennent conjointement de Morphonette et de DériF. La contribution originale de Morphonette concerne 1 388 relations, les 13 180 relations restantes étant fournies par DériF, dont 8 802 relations directes (entre un nom ou un adjectif et sa base verbale). L'apport original de DériF est donc de 4 378 relations indirectes.

La ligne 3 de la table 4 illustre le type de couples que l'on trouve dans Démonette, présents dans Morphonette et absents de DériF. Le plus souvent, ces couples mettent en jeu un ascendant verbal qui n'est pas ou n'est plus attesté en synchronie (*admonition*), ou dont le sens est partiellement démotivé (*acteur* ↔ *actrice* relativement au verbe *agir*) : dans tous les cas, les champs RELATION DIRECTE 1 et 2 sont vides.

À l'inverse, et en dehors des erreurs d'implémentation de DériF, les 4 378 relations indirectes qui en sont issus et qui sont absentes de Morphonette s'expliquent par les différences entre les approches implémentées dans ces deux ressources. Les critères de filtrage de Morphonette excluent en effet les regroupements les moins fréquents comme ceux en (20) :

- (20) abaissement/NOM ↔ abaisseur/NOM ; abattage/NOM ↔ abattement/NOM ;  
abolissement/NOM ↔ abolitif/ADJ ; abortion/NOM ↔ avorteur/NOM

La table 5 synthétise la distribution des relations indirectes extraites, en fonction des suffixes intervenant dans la formation des couples. Nous y distinguons les noms en *-eur* formés sur le radical savant de *X*, comme *agitateur*, étiquetés 'Xteur' des autres noms en *Xeur*, étiquetés 'Xeur', pour lesquels la base verbale est réalisée sous la forme du radical de l'imparfait, comme *régisseur*.

Xteur	158						
Xeuse	<b>3 938</b>	134					
Xrice	338	<b>978</b>	66				
Xment	<b>2 226</b>	98	<b>1 390</b>	40			
Xage	<b>2 946</b>	80	<b>1 932</b>	11	<b>1 656</b>		
Xion	<b>926</b>	<b>1 544</b>	208	<b>976</b>	320	240	
Xif	310	448	36	304	26	24	<b>918</b>
	Xeur	Xteur	Xeuse	Xrice	Xment	Xage	Xion

TABLE 5 : Distribution des couples, en fonction des suffixes. Les effectifs supérieurs à 900 couples sont en gras.

On remarque une disparité numérique dans la réalisation de ces relations.

1. Les plus représentées reflètent une attirance particulière du couple de suffixes pour un même type de radical de base : les suffixes *-ion*, *-if* et *-rice* (dits « savants ») préfèrent les radicaux savants des verbes, alors que les suffixes *-ment* et *-age* (« non savants ») tendent à sélectionner les radicaux de l'imparfait.
2. Les couples dont l'un des membres comporte un suffixe savant et l'autre un suffixe non savant ont une fréquence qui oscille entre 200 à 900. Les couples « savants » en Xteur/Xif et Xrice/Xif se trouvent également dans cette fourchette du fait du nombre relativement faible des dérivés en *-if* dans TLFnome (1 288).
3. D'autres schémas de couples mixtes instancient peu d'entrées car ils cumulent les handicaps évoqués ci-dessus.

Ces résultats mettent en lumière les combinaisons formelles les plus utilisées pour produire les relations morphologiques entre un nom d'agent masculin et son correspondant féminin, un nom d'agent et un nom d'activité, ou encore un nom d'agent et un adjectif de propriété. Ces hypothèses, qui doivent bien entendu être validées par une projection sur un corpus authentique, fournissent un indice de la productivité potentielle des suffixations en présence, et une prédiction de la structure du lexique en devenir.

## 4 Évaluation

Démonette pourrait à terme servir de ressource lexicale morphologique générale pour le français. Pour estimer la pertinence de son contenu actuel, nous l'avons comparé à Verbaction, une ressource morphologique utilisée régulièrement dans différentes tâches de TAL. Le but de la comparaison est de montrer que Démonette pourrait compléter ce lexique.



Verbaction, créé l'INaLF (aujourd'hui ATILF) puis à l'ERSS (aujourd'hui CLLE), se compose de 9 386 couples verbe-nom d'action morphologiquement apparentés. Les noms sont majoritairement construits par suffixation (*consommer* → *consommation*), mais on y trouve également beaucoup de conversions (*collection* → *collectionner* ; *copier* → *copie*). Verbaction a été construit de manière semi-automatique. La première version (Hathout et al., 2002) a été réalisée en effectuant une analyse basée sur un apprentissage de schémas de suffixation permettant de connecter les mots potentiellement apparentés morphologiquement. Les 5 058 couples verbe-nom obtenus ont ensuite été révisés manuellement par deux lexicographes. Verbaction a ensuite été complété par 4 335 couples en utilisant la boîte à outils Webaffix (Tanguy, Hathout, 2002) pour rechercher sur le Web des couples de mots dont l'un a la forme d'un nom déverbal et l'autre celle de son verbe de base. Les couples collectés ont aussi fait l'objet d'une révision manuelle.

La comparaison de Démonette avec Verbaction n'a porté que sur leur intersection potentielle, à savoir les couples verbe-nom d'action dont les deux membres appartiennent à TLFnome et dont les noms sont suffixés en *-age*, *-ion* ou *-ment*. 4 937 couples de Démonette et 5 313 couples de Verbaction vérifient ces conditions. Nous avons ensuite réalisé un calcul simple de rappel et de précision en utilisant Verbaction comme référence. Le rappel obtenu est de 84% et la précision de 90%. 479 couples verbe-nom présents dans Démonette sont absents de Verbaction, et, réciproquement, 1 351 couples de Verbaction manquent dans Démonette.

Sur un plan qualitatif, la comparaison de Démonette avec Verbaction fait principalement ressortir trois explications pour l'absence des 479 couples dans Verbaction:

1. Dans les cas où il existe une relation de parenté morphologique entre le verbe et le nom (*ébossier* → *ébossage* ; *stagner* → *stagnation* ; *ébahir* → *ébahissement*), l'absence du couple dans Verbaction est un oubli qu'il faut combler.
2. Dans certains couples, le nom dérivé du verbe est concurrencé pour différentes raisons par un autre nom d'action, plus fréquent, qui appartient à un domaine de spécialité (*ratrapper* → *rattrapement* ; *calibrer* → *calibration*) ou dénote l'un des participants à l'activité décrite par le verbe et non l'activité elle-même, par exemple, le moyen dans *embarquer* → *embarcation* ou *équiper* → *équipage*, le patient dans *vibrer* → *vibrion*.
3. Réciproquement, on trouve dans Démonette des couples erronés où le nom et le verbe ne sont pas apparentés (*évasion* # *évaser*), où le nom n'est pas déverbal (*outiller* # *outillage*), voire non construit en synchronie (*mentir* # *mention*, *munir* # *muniton*), ou encore où le verbe est archaïque (*aberrer* # *aberration*). La comparaison avec Verbaction nous a ainsi permis de répertorier un ensemble d'analyses à corriger (essentiellement dans DériF) et par suite dans Démonette. Ces cas de bruit avéré concernent 45 des 479 couples examinés (9%).

Pour les 1 351 couples présents dans Verbaction et absents de Démonette, on observe que 281 couples constituent de véritables cas de silence comme par exemple des noms en *-ion* dont la forme est construite sur le radical savant du verbe (*adjudger* → *adjudication*) et des noms en *-age* dont le radical du verbe de base est formellement très éloigné de celui utilisé en flexion (*paître* → *pacage*). D'autres absences s'expliquent par les choix opérés lors de la constitution de Démonette ou de Verbaction. Ont été en effet exclus de Démonette les 366 mots pour lesquels DériF propose plusieurs analyses comme *préchauffage* dont la base est soit *chauffage*, soit *préchauffer*. À l'inverse, Verbaction inclut des relations de conversion nom → verbe, d'où la présence de couples comme *prédilection* → *prédilectionner*. Plus intéressant, le codage de Verbaction comporte des couples qui exhibent une relation de sens indubitable, mais n'entretiennent pas de relation directe base → dérivé. Par exemple, avec *chromisation*, que Verbaction apparie à *chromer* (et, plus généralement, avec les 18 couples nom → verbe dont le nom en Xisation n'a pas de base Xiser enregistrée dans le TLF) on assiste à ce qui a été appelé dans (Dal 2004) puis (Dal, Namer, à paraître) un cas de relation morphologique ternaire schématisable par <X, Xiser, Xisation> et où la parenté entre Xiser et Xisation est non orientée : la construction de *chromisation*, sur le nom *chrome*, est consubstantielle, voire antérieure à celle de *chromiser*, absent du TLF.

## 5 Discussion

### 5.1 Usages en TAL

Les applications de TAL disposent pour le français de différentes ressources lexicales munies d'informations flexionnelles comme Morphalou (Romary et al., 2004) et Lexique (New et al., 2004), syntaxiques comme Lefff (Sagot et al., 2007) ou sémantiques comme EuroWordNet (Vossen, 1998) ou Wolf (Sagot, Fišer, 2008). Certaines sont spécialisées dans le traitement de données particulières comme les noms propres dans ProLexBase (Bouchou, Maurel, 2008). L'originalité de Démonette réside dans la variété des relations morphologiques codées, et dans les annotations morpho-sémantiques qui caractérisent les unités lexicales enregistrés dans la base. Celles-ci, tout d'abord, entretiennent entre elles des relations orientées multiples : une base est associée à plusieurs dérivés, les dérivés d'une même famille sont reliés entre eux. Ensuite, chaque mot comporte, une classe sémantique et autant de définitions, donc de paraphrases, qu'il entretient de relations directes ou indirectes avec son voisinage morphologique. Enfin, cette caractéristique concerne aussi les mots simples, puisque toutes les relations sont bi-orientées.

Ces trois propriétés sont exploitables dans différents types d'application. Comme l'a signalé Clark et al. (2008), l'amélioration de l'analyse du contenu textuel passe par l'enrichissement des bases par des annotations sémantiques. De plus, Démonette facilite les tâches de sélection des mots et d'identification du sens lexical. Comme avec WordNet et Wolf, ces facultés peuvent bénéficier à toutes les applications de compréhension ou d'interprétation automatique des langues (désambiguïsation sémantique, extraction d'information), ainsi qu'aux outils destinés à la production de contenu : donner la possibilité de choisir entre plusieurs mots, identifiables par leurs bases et sélectionnés à partir de leurs étiquettes sémantiques (agent, activité, propriété), voire de choisir une paraphrase pour les remplacer, est un avantage dont peuvent tirer parti les systèmes de normalisation, de résumé, de traduction assistée ou de génération de textes. Démonette, en association avec d'autres ressources comme des bases distributionnelles (Turney, Pantel, 2010), pourra également avoir d'autres utilisations comme la désambiguïsation et la substitution lexicale (McCarthy et al., 2004), la catégorisation sémantique des textes et la recherche d'information (Tsatsaronis, Panagiotopoulou, 2009).

## 5.2 Montée en puissance

La version actuelle de Démonette ne couvre qu'un fragment réduit du lexique du français et n'exploite qu'une petite partie des analyses de DériF et de Morphonette. Cette couverture sera augmentée progressivement pour intégrer l'ensemble de ces analyses, mais cette extension fera apparaître des difficultés nouvelles. Il faudra notamment augmenter le nombre des types sémantiques pour permettre l'ajout de nouveaux participants (le lieu, l'instrument, etc.), de nouvelles propriétés, etc. La sélection des relations indirectes à l'intérieur des familles dérivationnelles sera également un problème difficile. En l'absence d'études connues sur ces relations indirectes, nous proposons de les limiter en nous appuyant sur des critères comme la distance dans le graphe des relations directes entre les sommets connectés pondérée par une caractérisation statistique des séquences dérivationnelles. Certaines séquences comportant deux affixes successifs sont en effet si fréquentes qu'elles sont réinterprétées par les locuteurs comme de nouveaux affixes. C'est le cas de *-aliser* (Namer, 2013 ; Hathout, Namer, 2014), de *-isation* ou de *inXable* (Dal, Namer, à paraître).

Un second critère corrélé au précédent consiste à considérer qu'une relation indirecte « utile » est une relation pour laquelle les locuteurs peuvent formuler facilement une définition transversale. Les relations de ce type présentes dans la version actuelle de Démonette satisfont toutes à cette condition. En revanche, d'autres relations sont plus difficiles à définir d'une façon suffisamment régulière pour être intégrées à Démonette. C'est le cas pour les couples formés d'un nom masculin et d'un nom féminin en *-et:-ette* ou en *-ier:-ière* comme *cachet:cachette* ou *boulevardier:boulevardière*. La cachette (lieu où l'on cache quelque chose) n'est pas une version féminine du cachet (sceau) et la boulevardière (prostituée) n'exerce pas la même profession que le boulevardier (auteur de pièce de théâtre de boulevard). Les critères sur la distance dans le graphe et la possibilité d'une définition transversale ne se recouvrent donc pas totalement.

D'autres couples sont également difficilement définissables l'un relativement à l'autre, comme les adverbes en *-ment* et les verbes en *-iser* construits à partir de la même base adjectivale (21), ou encore les noms Xiste formés sur une base adjectivale X et les adjectifs déverbaux correspondant Xisable (22).

(21) stérilement/R:stériliser/V ; totalement/R:totaliser/V ; verbalement/R:verbaliser/V

(22) canonisable/A:canoniste/N ; conceptualisable/A:conceptualiste/N ; individualisable/A:individualiste/N

L'intégration des mots construits par conversion et par composition sera une autre source de difficultés. Si les définitions directes des dérivationnelles non affixales sont identiques à celles des affixations, celles des composés pourraient s'avérer plus délicates car il ne s'agit pas de relations binaires mais ternaires. Il faudra également prévoir une solution permettant de représenter les éléments de composition d'origine latine ou grecque.

Nous envisageons enfin l'intégration dans Démonette de relations morpho-sémantiques extraites de dictionnaires électroniques du domaine public en adaptant les techniques proposées par Hathout (2009).

## 6 Conclusion

Nous avons présenté dans cet article la première version de Démonette, une ressource lexicale comportant actuellement 31 204 relations annotées morphologiquement. Démonette est construite automatiquement à partir des résultats produits au moyen de deux approches que l'on oppose généralement l'une à l'autre en morphologie théorique : la première, Morphonette, exploite les analogies formelles entre lexèmes, et implémente les principes théoriques d'une morphologie réalisationnelle et paradigmaticque. La seconde, que réalise l'analyseur DériF, est basée sur l'application de règles linguistiques orientées, conçues et validées manuellement.

Nous avons mis en pratique, avec Démonette, une conception cumulative du sens des mots dérivés envisagé comme un empilement de propriétés sémantiques partielles induites par chacune des relations dérivationnelles dans lesquelles ce mot est impliqué. Démonette est en effet un réseau où les mots sont reliés entre eux par des relations directes ou

indirectes, selon la nature de leur parenté. Les relations dérivationnelles entre deux mots sont annotées par des définitions bi-orientées permettant de construire le sens de chaque mot par rapport à celui de l'autre. Un mot est alors décrit par autant de définitions qu'il entretient de relations avec d'autres unités dans le lexique. Les mots sont par ailleurs typés sémantiquement, permettant ainsi d'estimer la contribution de chaque catégorie sémantique dans l'élaboration du réseau morphologique tissé dans Démonette. En outre, les types sémantiques sont utilisés comme paramètres pour associer aux définitions concrètes des versions abstraites qui permettent de constituer les séries dérivationnelles.

La liste des dérivations actuellement représentées dans Démonette peut facilement être étendue aux adjectifs en *-oire* comme *exploratoire* aux noms en *-oir* comme *séchoir*, aux adjectifs en *-able* comme *calculable* ou aux noms de procès ou d'état en *-ure* et *-ance/-ence* comme *gravure* et *souffrance*, tous ces dérivés étant potentiellement connectés à une base verbale.

## Références

- BAAYEN, R. H., PIEPENBROCK, R., GULIKERS, L. (1995). The CELEX lexical database (release 2). Distributed by the Linguistic Data Consortium, University of Pennsylvania.
- BERNHARD D. (2006). Automatic Acquisition of Semantic Relationships from Morphological Relatedness. *FinTAL 2006*, Turku, Finland, Springer.
- BERNHARD D., CARTONI B., TRIBOUT D. (2011). A Task-Based Evaluation of French Morphological Resources and Tools. *Linguistic Issues in Language Technology* 5(2).
- BILGIN O., ÇETİNOĞLU Ö., OFLAZER K. (2004). Morphosemantic Relations In and Across Wornets. A study based on Turkish. Proceedings of GWC 2004. pp. 60-66. Brno, République Tchèque.
- BOUCHOU B., MAUREL D. (2008). Prolexbase et LMF : vers un standard pour les ressources lexicales sur les noms propres. *Traitement Automatique des Langues* 49(1), 61-88.
- CLARK P., FELLBAUM C., HOBBS J. R., HARRISON P., MURRAY B., THOMPSON J. (2008). Augmenting WordNet for deep understanding of text. *Proceedings of Semantics in Text Processing*, Venezia, ACL.
- CREUTZ M., LAGUS K. (2005). *Unsupervised morpheme segmentation and morphology induction from text corpora using Morfessor 1.0*. Tech. Rep. A81, Helsinki University of Technology.
- DAL G. (2004). *Vers une morphologie de l'évidence : d'une morphologie de l'input à une morphologie de l'output*. Villeneuve d'Ascq, Université de Lille3. Mémoire d'Habilitation à Diriger les Recherches.
- DAL G., HATHOUT N., NAMER F. (1999). Construire un lexique dérivationnel: théorie et réalisations. *TALN-1999* Cargèse, Université Paris 7.
- DAL G., NAMER F. (à paraître). La fréquence en morphologie : pour quels usages ?, *Langages*.
- FELLBAUM C., MILLER G. A. (2003). Morphosemantic Links in WordNet. *Traitement Automatique des Langues* 44(2):69-80.
- FRADIN B. (2003). *Nouvelles approches en morphologie*, Paris, PUF.
- GOLDSMITH J. (2001). Unsupervised learning of the morphology of natural language. *Computational Linguistics* 27(2):153-198.
- HARRIS Z. (1955). From Phoneme to Morpheme. *Language* 31(2), 190-222.
- HATHOUT N. (2009). Acquisition of morphological families and derivational series from a machine readable dictionary. *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*. F. Montermini, G. Boyé, J. Tseng. Cambridge, Mass., Cascadilla Proceedings Project: 166-180.
- HATHOUT N. (2011). Morphonette: a paradigm-based morphological network. *Lingue e linguaggio* 2011(2), 245-264.
- HATHOUT, N., NAMER, F. ET DAL, G. (2002). An Experimental Constructional Database: The MorTAL Project. In Boucher, P. éditeur, *Many Morphologies*. Cascadilla, Somerville, Mass.
- HATHOUT N., NAMER F. (2011). Règles et paradigmes en morphologie informatique lexématique. *TALN-2011*, Montpellier, LIRMM/ATALA.
- HATHOUT N., NAMER F. (2014). Discrepancy between form and meaning in Word Formation: the case of over- and under-marking in French. *Morphology and meaning (Selected papers from the 15th International Morphology Meeting, Vienna, February 2010)* F. Rainer, W. U. Dressler, F. Gardani, H. C. Luschützky. Amsterdam, John Benjamins, 177 – 190.

- KERLEROUX F. (2008). Des noms indistincts. *La raison morphologique. Hommage à la mémoire de Danielle Corbin*. B. Fradin. Amsterdam / Philadelphia, John Benjamins, 113-132.
- LAFOURCADE M., JOUBERT A., 2013. Bénéfices et limites de l'acquisition lexicale dans l'expérience JeuxDeMots. In: Gala N., Zock M. (dir.), *Ressources Lexicales: Contenu, construction, utilisation, évaluation*, pp. 187-216. *Linguisticae Investigationes, Supplementa 30*, John Benjamins.
- LEPAGE Y. (1998). Solving analogies on words: an algorithm. *17th international conference on Computational Linguistics*.
- MCCARTHY D., KOELING R., WEEDS J., CARROLL J. (2004). Finding predominant senses in untagged text. *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pp. 577-583. Barcelone.
- NAMER F. (2009). *Morphologie, Lexique et TAL : l'analyseur DériF*. London, Hermes Sciences Publishing.
- NAMER F. (2013). A Rule-Based Morphosemantic Analyzer for French for a Fine-Grained Semantic Annotation of Texts. *SFCM 2013*. C. Mahlow, M. Piotrowski. Heidelberg, Springer: 93-115.
- NEW B., PALLIER C., BRYBAERT M., FERRAND L. (2004). Lexique 2 : A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers* 36(3), 516-524.
- PALA K., HLAVÁČKOVÁ D. (2007). Derivational Relations in Czech WordNet . *Balto-Slavonic Natural Language Processing 2007* , pp 75–81, Prague . République Tchèque.
- ROCHÉ M., BOYÉ G., HATHOUT N., LIGNON S., PLÉNAT M. (2011). *Des Unités Morphologiques au Lexique*. Paris, Hermès.
- ROMARY L., SALMON-ALT S., FRANCOPOULO G. (2004). Standards going concrete : from LMF to Morphalou. *COLING*, Genève.
- RUPPENHOFER J., BAKER C. F., FILLMORE C. J. (2002). The FrameNet Database and Software Tools. *Proceedings of the Tenth Euralex International Congress*. pp. 271-375. Copenhagen, Danemark.
- SAGOT B., FIŠER S. (2008). Construction d'un wordnet libre du français à partir de ressources multilingues. *TALN-2008*, Avignon, ATALA.
- STROPPA N., YVON F. (2006). Du quatrième de proportion comme principe inductif : une proposition et son application à l'apprentissage de la morphologie. *TAL* 47(1), 33-59.
- TANGUY L., HATHOUT N. (2002). Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du Web. *TALN-2002*, Nancy, ATALA.
- TSATSARONIS, G., PANAGIOTOPOULOU, V. (2009). A generalized vector space model for text retrieval based on semantic relatedness. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, pp. 70-78. Athènes. Grèce.
- TURNER, P. D., PANTEL, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1), 141-188.
- VOSSEN P. ed. (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, . Dordrecht, Kluwer Academic Publishers.
- WATTS D. J., STROGATZ S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440-442.