

Représentation sémantique des familles dérivationnelles au moyen de frames morphosémantiques

Daniele Sanacore¹ Nabil Hathout¹ Fiammetta Namer²

(1) CLLE, CNRS & Université de Toulouse

(2) ATILF, CNRS & Université de Lorraine

RÉSUMÉ

L'article présente un formalisme de représentation des relations morphologiques dérivationnelles inspiré de la Sémantique des Frames. La description morphosémantique y est réalisée au niveau des familles dérivationnelles au moyen de frames morphosémantiques dans lesquels les lexèmes sont définis les uns relativement aux autres. Les frames morphosémantiques permettent de rendre compte de la structure paradigmatique du lexique morphologique par l'alignement des familles qui présentent les mêmes oppositions de sens. La seconde partie de l'article est consacrée aux données qui seront utilisées pour produire (semi-) automatiquement ces représentations.

ABSTRACT

Semantic representation of derivational families by means of morphosemantic frames

In this paper, we propose a formalism for the morphosemantic description of the morphological families inspired by Frame Semantics. In this formalism, the lexemes are typed semantically and ontologically and are defined with respect to the other members of their family. The morphosemantic frames describe the semantic structure of sets of morphological families. Having identical semantic structures, these families can be aligned into derivational paradigms. In the last section of the paper, we review the data that will be used to (semi-) automatically generate these frames.

MOTS-CLÉS : morphologie dérivationnelle, morphologie paradigmatique, sémantique des frames, types morphosémantiques, types ontologiques.

KEYWORDS: derivational morphology, paradigmatic morphology, frame semantics, morphosemantic types, ontological types.

1 Introduction

Nous proposons dans cet article un nouveau formalisme pour la représentation sémantique des relations dérivationnelles.¹ Ce travail est une contribution au développement de Démonette (Hathout & Namer, 2014, 2016; Namer *et al.*, 2019) une ressource morphologique dérivationnelle du français. Les entrées de Démonette sont des relations dérivationnelles entre des lexèmes de la même famille dérivationnelle. Ces relations peuvent être directes (*rédiger - rédacteur*) ou indirectes (*rédaction - rédacteur*). Nous considérons dans Démonette que le sens d'un lexème construit est déterminé par l'ensemble des relations sémantiques dans lesquelles il est impliqué. Le formalisme proposé, que nous appellerons **frames morphosémantiques**, permet donc de décrire l'ensemble des relations

1. Ce travail bénéficie du soutien de l'ANR 17-CE23-0005.

morphosémantiques qui existent entre les membres d’une famille dérivationnelle, mais aussi leurs relations argumentales, la et la catégorie ontologique des lexèmes. Les relations morphosémantiques sont décrites au moyen de **gloses morphosémantiques** qui définissent le sens des membres de chaque famille les uns relativement aux autres.

2 Fondements théoriques

La morphologie dérivationnelle abandonne progressivement le morphème comme unité d’analyse en faveur du lexème (Aronoff, 1976; Anderson, 1992; Fradin, 2003). Le morphème est en effet trop rigide et ne permet pas de décrire les très nombreux décalages entre forme et sens, comme dans le cas des constructions dites parasyntétiques (*continent* → *intercontinental*) (Hathout & Namer, 2018; Namer & Hathout, 2019). Un autre développement plus récent est l’adoption d’approches paradigmatiques qui rendent compte d’une manière plus complète des très nombreuses régularités qui existent dans le lexique dérivationnel (Van Marle, 1984; Stump, 1991; Bauer, 1997; Booij, 2008, parmi d’autres). C’est dans ce cadre que s’inscrit notre proposition. Les paradigmes dérivationnels sont basés sur trois structures : les familles, les séries et les paradigmes. Elles sont illustrées en Figure 1.

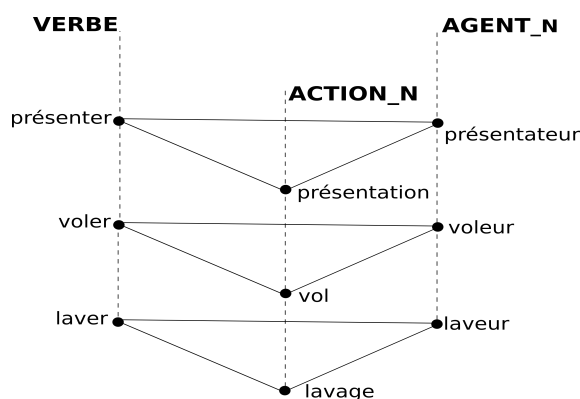


FIGURE 1 – Extrait d’un paradigme dérivationnel

Une **famille dérivationnelle** est un ensemble de lexèmes connectés par des relations dérivationnelles directes ou indirectes (Hathout, 2011). Une relation dérivationnelle directe connecte un dérivé à sa base (*laver* - *lavage*) ou vice-versa. Les autres relations dérivationnelles sont indirectes comme *lavage* - *laveur* qui dérivent tous les deux de *laver*. Une **série dérivationnelle** se compose de lexèmes qui se trouvent dans les mêmes relations de contraste avec les autres membres de leurs familles. Par exemple, *coiffage*, *tranchage* et *calage* appartient à la série composée des noms d’action en *-age*. Un **paradigme dérivationnel** est un empilement de **sous-familles** (c’est-à-dire de sous-ensembles de familles dérivationnelles) dont les membres sont liés par les mêmes relations, et qui de ce fait sont superposables. Dans cet article, nous nous intéressons aux relations sémantiques qui s’établissent dans les familles et les paradigmes dérivationnels et notamment à la manière dont la caractérisation sémantique des familles morphologiques met en évidence l’organisation paradigmatique du lexique dérivationnel.

3 Démonette

Démonette est une base de données morphologique dérivationnelle alimentée par des ressources existantes de nature variée (Namer *et al.*, 2019). Dans cette ressource, chaque relation dérivationnelle détermine un ensemble de propriétés sémantiques des lexèmes qu'elle connecte. Le sens d'un dérivé est conçu comme la combinaison des propriétés sémantiques induites par toutes les relations dérivationnelles dans lesquelles il est impliqué. Une entrée dans Démonette décrit un couple de lexèmes d'une famille dérivationnelle, Lex_1 et Lex_2 . Elle fournit notamment les propriétés formelles, morphologiques et sémantiques des deux lexèmes et les propriétés formelles, morphologiques et sémantiques de leur relation. La description sémantique des lexèmes et des relations est illustrée dans la Table 1. Elle se compose d'une description ontologique de chacun des deux lexèmes (Typ_1 et Typ_2) et d'une caractérisation sémantique de la relation entre Lex_1 et Lex_2 (trois dernières colonnes).

Lex ₁ - Lex ₂	Typ ₁	Typ ₂	Relation sémantique		
			Type général	Sém. Lex ₁	Schéma Lex ₁
danseur _N - danser _V	person	sit.dyn	entité-situation	agent	ce(lui) qui Lex ₂
admirer _V - admirateur _N	sit.stat	person	situation-entité	jugement	ressentir ce que ressent Lex ₂

TABLE 1: Description sémantique des couples de lexèmes dans Démonette

4 La sémantique des frames

Le sémantique des frames est une théorie de la représentation des connaissances basée sur les frames. Un frame est une structure conceptuelle qui décrit une situation, un objet ou un évènement ainsi que ses participants et leurs propriétés. Les participants d'un frame, appelés **frame elements (FE)**, sont décrits par des rôles sémantiques. Il existe plusieurs implémentations de la sémantique des frames comme FrameNet (Fillmore, 2006; Johnson *et al.*, 2003; Ruppenhofer *et al.*, 2006) pour l'anglais ou Asfalda (Candito *et al.*, 2014) pour le français.

Un frame se compose (i) d'une glose globale qui décrit la situation représentée et la manière dont les différents FE interagissent à l'intérieur de celle-ci; (ii) des gloses partielles qui décrivent la situation relativement à chacun des FE; (iii) d'un ensemble d'unités lexicales qui évoquent le frame; (iv) d'un ensemble de phrases qui réalisent le frame. Par exemple, dans Asfalda, le frame CONTACTING est défini par la glose globale en anglais (1); le FE **Addressee** est décrit par la glose partielle (2); la liste des unités lexicales du français qui évoquent la situation exprimée par le frame CONTACTING est donnée en (3); la phrase (4) est un exemple de réalisation de ce frame.

- (1) CONTACTING : A **Communicator** directs a communication to an **Addressee** and/or at a particular **Address**, for a particular **Purpose** and/or about a particular **Topic**. A **Location_of_communicator** can be expressed.
- (2) **Addressee** : The person that receives the message from the **Communicator**.
- (3) *appel téléphonique.n, appel.n, appeler.v, contact.n, contacter.v, coup de fil.n, coup de téléphone.n, écrire.v, joindre.v, téléphoner.v, toucher.v*
- (4) FR : *Faisant contre mauvaise fortune bon coeur, l' amiral Higgins a alors **appelé** au téléphone les journalistes britanniques pour les délier de leurs engagements*

5 Adaptation des frames sémantiques à la morphosémantique

Les correspondances qui existent entre les frames sémantiques et les paradigmes dérivationnels permettent une adaptation immédiate des premiers pour décrire les seconds. En effet, les unités lexicales qui évoquent les frames correspondent aux (sous-) familles qui composent les paradigmes. Les FE correspondent aux séries dérivationnelles, c'est-à-dire à des lexèmes jouent le même rôle dans leurs familles. La description du frame correspond au schéma du paradigme dérivationnel, c'est-à-dire à la description des relations qui existent entre les séries. Cette description peut se faire au moyen d'un ensemble de gloses morphosémantiques. Schématiquement, la correspondance entre les frames et les paradigmes peut être résumée comme suit :

frame sémantique	paradigme
description du frame	schéma du paradigme
frame element	série de lexèmes
unités lexicales du frame	familles dérivationnelles

La description des FE comporte trois niveaux : le niveau relationnel réunit les **gloses morphosémantiques** ; le niveau argumental décrit les rôles des membres des séries au sein de leur familles ; le niveau ontologique donne la catégorie des membres de ces séries. La dimension ontologique est indépendante de la dimension argumentale. En revanche, les dimensions relationnelles et argumentales sont interdépendantes tout en étant distinctes. Pour présenter plus en détail ces trois niveaux de description, considérons la sous-famille dérivationnelle (5).

(5) *laver* _V ; *lavage* _N ; *lavoir* _N ; *laverie* _N ; *laveur* _N ; *laveuse* _N ; *lavette* _N ; *lavable* _A

Au **niveau relationnel**, on trouve les gloses morphosémantiques qui décrivent les relations qui connectent les lexèmes : les éléments d'une famille y sont définis mutuellement au moyen d'énoncés non orientés comme en (6), de sorte que la définition globale de la famille est composée de l'union de toutes ces gloses simples. Notons que plusieurs membres d'une famille sont susceptibles d'occuper la même position dans une glose de la même manière qu'ils sont susceptibles d'occuper la même case dans le paradigme. C'est le cas de *laveur* et *laveuse* en (6a), puisque les deux noms sont les agents de *laver*. De même, *lavoir* et *laverie* en (6c) décrivent des lieux scéniques où se déroule l'activité *laver*.

- (6) a. Un **laveur** (une **laveuse**) **lave** quelque chose
 b. Quelque chose est **lavable** si on peut la **laver**
 c. On **lave** quelque chose dans un **lavoir** (une **laverie**)

Pour obtenir des gloses morphosémantiques intelligibles et suffisamment naturelles, le nombre d'éléments dans chaque glose est limité à deux ou trois. De ce fait, il n'est pas envisageable de réunir l'ensemble des lexèmes de la famille dans une glose globale unique. Une glose morphosémantique associe en réalité deux ou plusieurs séries dérivationnelles et décrit une superposition de couples au sein des familles du paradigme. Les couples qui instancient la glose se trouvent tous dans la même relation comme l'illustre la Table 2.

Au **niveau argumental**, un rôle sémantique est attribué à chaque membre de la (sous-) famille (ou plus exactement à chaque FE instancié par ce membre), en fonction de la position que le lexème occupe dans la structure argumentale des prédicats de ses gloses du niveau relationnel. Par exemple, pour le frame qui contient la famille de *laver* (5), les rôles sémantiques des FE instanciés par les membres de cette famille sont donnés en Table 3.

	PRÉD.VSUP		PRÉDICAT	
On réalise un(e)	lavage	quand on	lave	qqc
On réalise un(e)	vol	quand on	vole	qqc
On réalise un(e)	gonflement	quand on	gonfle	qqc
On réalise un(e)	présentation	quand on	présente	qqc
On réalise un(e)	défense	quand on	défend	qqc
On réalise un(e)	découverte	quand on	découvre	qqc

TABLE 2: Instances d’une glose morphosémantique qui relie deux séries dérivationnelles. La première est composée de noms d’action et la seconde des verbes correspondants

FE	rôle	FE	rôle
laver	prédicat	lavoir / laverie	lieu
lavage	prédicat à verbe support (pratiquer/réaliser)	lavette	instrument
laveur / laveuse	agent	lavable	potentialité

TABLE 3: Niveau argumental des FE du frame morphosémantique qui contient la famille de *laver*

Au **niveau ontologique**, nous utilisons pour les noms la version du projet *Fr-SemCor* (Barque *et al.*, 2020) des *Unique Beginners for Nouns* (UB) de *WordNet* (Miller *et al.*, 1990). Cette ontologie est donnée en Table 4. Son niveau de granularité relativement fin permet la caractérisation des FE qui participent à certaines relations dérivationnelles spécifiques comme la suffixation en *-aie* qui construit des noms de plantations à partir de noms de plantes (*oranger* → *orangerie*; *palmier* → *palmeraie*). Cette ontologie est complétée par une catégorie SITUATION pour les verbes et une catégorie MODIFIER pour les adjectifs. Nous n’avons pas utilisé les 15 premières catégories de verbes de WordNet (*bodily function and care, change, communication, competition, etc.*) comme nous l’avons fait pour les noms parce que les contraintes imposées par les procédés dérivationnels qui mettent en jeu des verbes sont très générales.

Entity			Situation	
Animate Entity	Non Animate Entity	Abstract Entity	Stative Situation	Dynamic situation
animal	artifact	cognition	attribute	act
person	groupxartifact	groupxcognition	state	event
groupxanimal	food	communication	feeling	
groupxperson	substance	group		
	object	part		
	plant	quantity		
	body	possession		
		relation		
		phenomenon		

TABLE 4: Catégories ontologiques utilisées dans le projet *Fr-SemCor*

Un frame morphosémantique se compose d’un ensemble de gloses qui définissent ses FE les uns relativement aux autres, d’une caractérisation ontologique et argumentale des FE et d’un ensemble de familles qui instancient ces FE. La Figure 2 présente une partie du frame morphosémantique qui contient la famille de *laver* (5). Ce frame réunit trois FE (prédicat verbal, prédicat nominal,

nom d’agent) définit mutuellement par 3 gloses morphosémantiques pour l’une des familles, en l’occurrence celle de *laver*. Par exemple, la première glose décrit la relation qui existe entre le prédicat nominal exemplifié par *lavage* et le prédicat verbal exemplifié par *laver*. Les catégories grammaticales et ontologiques et les rôles de ces FE sont donnés dans la seconde partie du frame. La dernière partie du frame liste les familles qui composent le frame, en l’occurrence, celles de *laver*, de *voler* et de *présenter*. Chaque membre de chaque famille est assigné à un FE. Par exemple, *voleur* et *voleuse* sont des instances de FE3.

Gloses MorphoSémantiques

On réalise un(e) **lavage**=FE2 quand on **lave**=FE1 qqc

Un(e) **laveur**=FE3 réalise un(e) **lavage**=FE2

Un(e) **laveur**=FE3 **lave**=FE1 qqc

	PoS.Onto	rôle
FE1	V.SITUATION	PRÉDICAT
FE2	N.ACTIVITY	PRED.VSUP
FE3	N.PERSON	AGENT

FE1	FE2	FE3
laver	lavage	laveur/laveuse
voler	vol	voleur/voleuse
présenter	présentation	présentateur/présentatrice

FIGURE 2 – Partie du frame morphosémantique qui contient la famille de *laver*

6 Génération (semi-) automatique des frames morphosémantiques

Nous venons de voir dans ce qui précède la motivation et les principes qui gouvernent la conception de frames morphosémantiques. Différentes ressources existantes peuvent être utilisées pour les produire (semi-)automatiquement pour un ensemble significatif de familles et de paradigmes de Démonette. Une première ressource est Glawinette, un lexique de familles et de séries dérivationnelles du français (Hathout *et al.*, 2020) qui décrit 160113 couples de lexèmes regroupés en 15895 familles. Les couples de lexèmes sont caractérisés par des couples de schémas dérivationnels qui permettent de déduire les FE auxquels ils peuvent se rattacher dans un frame morphosémantique. Nous envisageons d’utiliser les définitions lexicographiques de dictionnaires comme GLAWI (Sajous & Hathout, 2015) pour la génération de gloses morphosémantiques standardisées et la détermination des rôles argumentaux et des catégories ontologiques.

Malgré les variations qu’elles peuvent présenter, les définitions sont une source de connaissance riche et adaptée à la production de frames morphosémantiques comme l’illustrent les exemples sous (7). Les définitions des lexèmes dérivés contiennent souvent un autre membre de leurs famille dérivationnelle ; ces mots sont soulignés dans les exemples de (7) ; ces éléments sont fournis par Glawinette ou peuvent à défaut être identifiés sur la base des analogies qu’ils permettent de former. Le genre prochain dans les définitions permet par ailleurs d’identifier la catégorie ontologique de l’entrée ; ces genres sont encadrés dans les exemples de (7) ; leur extraction peut se faire directement à partir de GLAWI qui

fournit des analyses syntaxiques en dépendance de toutes les définitions. Une fois le genre prochain identifié, la catégorie ontologique peut être déduite en utilisant des correspondances comme en (8).

- (7) a. OBSERVATEUR : personne *qui observe*
b. SOIGNEUR : personne *qui donne des soins et s'occupe de l'état physique d'un sportif*
c. OBSERVATOIRE : lieu *d'où l'on peut observer, surveiller l'ennemi*
d. DÉMAIGRISSEMENT : action *de démaigrir*
e. ÉTIQUETEUSE : machine *qui fait l'étiquetage*

- (8) *personne* → person, human being
lieu → artifact OR location
action → act, activity
machine → artifact

La structure syntaxique de la définition peut aussi servir de matrice pour la génération des gloses morphosémantiques. Par exemple, la définition donnée dans la première ligne des des exemples en (9) permet de produire la glose de la 2^e ligne et de déduire les rôles et les catégories ontologiques des FE listées dans la 3^e ligne. Le nombre des paradigmes du français permet de réaliser cette tâche au moyen de patrons. Elle peut également être réalisées au moyen de réseaux de neurones en adaptant la méthode proposée par (Mickus *et al.*, 2020).

- (9) a. DONATEUR : *personne qui a fait une donation*
*Un **donateur** fait une **donation***
AGENT = **donateur** ; PRÉDICAT NOMINAL = **donation** ; VERBE SUPPORT = **faire**
- b. EXPOSANT : *personne qui expose un fait*
*Un **exposant** expose quelque chose*
AGENT = **exposant** ; PRÉDICAT = **exposer**
- c. OBSERVABLE : *qui peut être observé*
*On peut **observer** quelque chose ou quelqu'un qui est **observable***
MODIFIEUR = **observable** ; PRÉDICAT = **observer**

7 Conclusion

Nous avons montré dans cet article comment des structures inspirées de la sémantique des frames peuvent être adaptées pour permettre la description des relations morphosémantiques qui s'établissent dans les familles dérivationnelles. Les frames morphosémantiques rendent compte également de la superposition de familles et de ces relations à l'intérieur des paradigmes dérivationnelles. Les frames morphosémantiques peuvent être produits de manière (semi-)automatique à partir de données structurées provenant des dictionnaires, comme GLAWI et Glawinette.

Références

- ANDERSON S. R. (1992). *A-morphous morphology*, volume 62. Cambridge University Press.
- ARONOFF M. (1976). *Word Formation in Generative Grammar*. Linguistic Inquiry Monographs. Cambridge, MA : MIT Press.
- BARQUE L., HAAS P., HUYGHE R., TRIBOUT D., CANDITO M., CRABBÉ B. & SEGONNE V. (2020). FrSemCor : Annotating a French corpus with supersenses. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*.
- BAUER L. (1997). Derivational paradigms. In *Yearbook of Morphology 1996*, p. 243–256. Springer.
- BOOIJ G. (2008). Paradigmatic morphology. *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, p. 29–38.
- CANDITO M., AMSILI P., BARQUE L., BENAMARA ZITOUNE F., DE CHALENDAR G., DJEMAA M., HAAS P., HUYGHE R., YANNICK MATHIEU Y., MULLER P., SAGOT B. & VIEU L. (2014). Developing a French Framenet : Methodology and first results. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland.
- FILLMORE C. J. E. A. (2006). Frame semantics. *Cognitive linguistics : Basic readings*, **34**, 373–400.
- FRADIN B. (2003). *Nouvelles approches en morphologie*. Presses Universitaires de France.
- HATHOUT N. (2011). Morphonette : a paradigm-based morphological network. *Lingue e linguaggio*, **10**(2), 245–264.
- HATHOUT N. & NAMER F. (2014). Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology*, **11**(5), 125–168.
- HATHOUT N. & NAMER F. (2016). Giving lexical resources a second life : Démonette, a multi-sourced morpho-semantic network for french. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.
- HATHOUT N. & NAMER F. (2018). La parasynthèse à travers les modèles : des RCL au ParaDis. In O. BONAMI, G. BOYÉ, G. DAL, H. GIRAUDO & F. NAMER, Éd., *The lexeme in descriptive and theoretical morphology*, p. 365–399. Langage Sciences Press.
- HATHOUT N., SAJOUS F., CALDERONE B. & NAMER F. (2020). Glawinette : a linguistically motivated derivational description of French acquired from GLAWI. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*.
- JOHNSON C. R., SCHWARZER-PETRUCK M., BAKER C. F., ELLSWORTH M., RUPPENHOFER J. & FILLMORE C. J. (2003). *Framenet : Theory and practice*. Rapport interne, International Computer Science Institute, Berkeley, CA.
- MICKUS T., CONSTANT M. & PAPERNO D. (2020). Génération automatique de définition pour le français. In *Actes de la 27^e conférence annuelle sur le traitement automatique des langues naturelles (TALN-2020)*.
- MILLER G. A., BECKWITH R., FELLBAUM C., GROSS D. & MILLER K. J. (1990). Introduction to WordNet : An on-line lexical database. *International Journal of Lexicography*, **3**(4), 335–391.
- NAMER F., BARQUE L., BONAMI O., HAAS P., HATHOUT N. & TRIBOUT D. (2019). Démonette2 – une base de données dérivationnelles du français à grande échelle : premiers résultats. In *Actes de la 26^e conférence annuelle sur le traitement automatique des langues naturelles (TALN-2019)*, p. 233–243, Toulouse.

- NAMER F. & HATHOUT N. (2019). ParaDis and Démonette : From theory to resources for derivational paradigms. In *Proceedings of the Second International Workshop on Resources and Tools for Derivational Morphology*, p. 5–14.
- RUPPENHOFER J., ELLSWORTH M., PETRUCK M. R., JOHNSON C. R. & SCHEFFCZYK J. (2006). *FrameNet II : Extended Theory and Practice*. Berkeley, California : International Computer Science Institute. Distributed with the FrameNet data.
- SAJOUS F. & HATHOUT N. (2015). GLAWI, a free XML-encoded Machine-Readable Dictionary built from the French Wiktionary. In *Proceedings of the of the eLex 2015 conference*, p. 405–426, Herstmonceux, England.
- STUMP G. T. (1991). A paradigm-based theory of morphosemantic mismatches. *Language*, p. 675–725.
- VAN MARLE J. (1984). *On the paradigmatic dimension of morphological creativity*. Dordrecht : foris.