Semantic descriptions of French derivational relations in a families-and-paradigms framework

Daniele Sanacore
CLLE, CNRS
University of Toulouse
daniele.sanacore@univ-tlse2.fr

Nabil Hathout
CLLE, CNRS
University of Toulouse
Nabil.Hathout@univ-tlse2.fr

Fiammetta Namer
ATILF, CNRS
University of Lorraine
Fiammetta.Namer@univ-lorraine.fr

Abstract

This paper proposes a new way to represent morphosemantic regularities in derivational paradigms of French in the context of derivational morphology. Starting from what has already been done in Démonette, a derivational morphological lexical resource for French, we show how structures inspired by Frame Semantics and FrameNet could help with the problem of the efficient representation of morphosemantic regularities in derivational paradigms. This first phase of the experiment consisted in the representation of four French derivational subfamilies of the French lexicon with a frame-like structure in order to show how this approach could work.

1 Introduction

An increasing number of lexical resources containing word formation descriptions are currently developed for many languages. If we start from the basic assumption that morphology is the study of systematic covariation in the form and meaning of words (Haspelmath and Sims, 2013), one problem that remains unsolved in the context of derivational morphology is finding an efficient way to represent morphosemantic regularities that are present in the derivational lexicon. In this paper, we address this issue in the framework of paradigmatic morphology. The objective is to describe the morphosemantic relations contained in the lexicon and design semantic representations compatible with morphological resources that could be used in NLP and experimental linguistics. Starting from what has already been done with *Démonette* (Hathout and Namer, 2014, 2016), we propose a representation of paradigmatic regularities in the lexicon by using structures inspired by Frame Semantics (Fillmore et al., 2006) and used in resources like *FrameNet* (Baker et al., 1998). Although differences exist between the objectives of *FrameNet* (document the range of semantic and syntactic combinatory possibilities of each word in each of its senses through objects called "frames") and *Démonette* (representing morphological regularities in the lexicon) frame-like structures could help us achieve our objective.

2 Definitions

Derivational families. A derivational family is a set of lexemes connected by morphological derivational relations (Hathout, 2009). This extensive definition includes also forms with suppletive stems (*hippodrome* 'racecourse' in the family of *cheval* 'horse'). An example of derivational family for French is the one built around the verb *laver* 'to wash' in (1):

(1) *laver* 'to wash'; *lavage* 'washing'; *lavoir* 'wash house'; *laverie* 'laundromat'; *laveur* 'washer (male)'; *laveuse* 'washer (female)'; *lavette* 'dishcloth'; *lavable* 'washable'; *lavement* 'enema'

In derivational families we can find two types of derivational relations between lexemes: direct relations and indirect relations. A **direct derivational relation** connects a lexeme directly

with one of its descendants or ascendants in the derivational family, for instance *laver* 'to wash' and *laveur* 'washer (male)'. On the other hand, an **indirect derivational relation** connects more distant elements of the family, e.g *laveur* 'washer (male)' with *lavage* 'washing'. In fact, both *laveur* and *lavage* are derived from *laver*.

Paradigmatic systems. A paradigmatic system is a collection of (partial) families that are aligned in terms of the content-based relations that their members entertain (Bonami and Strnadová, 2018). The CONTENT is the specification of the syntactic and semantic properties of a word, while the FORM is the specification of its phonology e/o orthography. The notion of paradigmatic system can be used both for inflectional and derivational morphology. An example of paradigmatic system can be illustrated by the following four subfamilies for the verbs *imprimer* 'to print', *souder* 'to weld', *laver* 'to wash' and *nettoyer* 'to clean':

(2)

verb	**agent_m**	**adj**	**action noun**
imprimer	imprimeur	imprimable	impression
souder	soudeur	soudable	soudage
laver	laveur	lavable	lavage
nettoyer	nettoyeur	nettoyable	nettoyage

The derivational relation between the verb *imprimer* and the masculine human agent noun *imprimeur* is the same as the derivational relations that link *souder* and *soudeur*, *laver* with *laveur* and *nettoyer* with *nettoyeur*. Another alignment can be found between the relations connecting the verbs with the derived modal adjective: the derivational relation between *imprimer* and *imprimable* is the same as the relation between *souder* and *soudable*, *laver* and *lavable* and *nettoyer* and *nettoyable*. Ultimately, a third alignment can be seen about the relations linking the verb (*imprimer, souder, laver, nettoyer*) with the respective action nouns (*impression, soudage, lavage, nettoyage*). It is important here to specify that the notion of alignment is based on content, rather than form. Pairs of words are aligned if they contrast in the same way. When an alignment of same derivational relations between couples of lexemes is found, these relations compose a **derivational series**.

3 Démonette

The problem of organizing morphosemantic description has been approached by resources like *Démonette* (Hathout et al., 2017; Hathout and Namer, 2014), a French derivational database. *Démonette* is a resource designed for the description of word formation in French. Its construction is based on the fundamental assumption that morphology is relational and each relation where a given word is involved contributes to its meaning. *Démonette* seeks a complete, redundant and explicit description of all the properties of each relation and each description of a relation is independent from the others. For this reason, entries in the *Démonette* database do not describe the properties of the derivatives, they describe instead properties of the derivational relations connecting two lexemes. Entries are thus pairs of morphologically related words (w_1, w_2) belonging to the same derivational family, such as *laver* $\rightarrow$ *laveur*.

Relations in *Démonette* are characterized by their orientation. *Démonette* is a directed graph where a relation $(w_1 \leftarrow w_2)$ describes the morphological motivation of w_1 with respect to w_2. Most of the lexemes are connected with each other in both directions. (Hathout and Namer, 2016). Direct relations in *Démonette* may be descending or ascending: the first connect a derived lexeme to its base or to a more distant ascendant (*laver* $\leftarrow$ *laveur*) while the latter connect a lexeme to its derivative or to a more distant descendant (*laveur* $\leftarrow$ *laver*).

Among the existing fields used to describe derivational relations in the *Démonette* database, an important role is played by the four fields used for the semantic description. Currently, there are two fields expressing the semantic type of w_1 and w_2, one for the concrete definition giving

the meaning of w_1 with respect to w_2 and one for the abstract definition where w_2 is replaced by its semantic type. Abstract definitions are important to highlight morphosemantic paradigms in the database. In fact, relations with the same abstract definition highlight regularities in the lexicon and form a derivational series, as in Table 1:

W1	W2	Type W1	Type W2	Concrete definition	Abstract definition
laveuse	laver	@AGF	@	"she who performs the action of laver "	"she who performs the action of @"
nettoyeuse	nettoyer	@AGF	@	"she who performs the action of nettoyer"	"she who performs the action of @"
imprimeuse	imprimer	@AGF	@	"she who performs the action of imprimer"	"she who performs the action of @"

Table 1: Semantic types, concrete and abstract definitions

For what concerns the semantic typing provided in Table 1, (e.g. @AGF for *laveuse, nettoyeuse* and *imprimeuse*), the problem is that it actually merges two levels of morphosemantic information: the ontological category of the described lexeme and its semantic role. Given that the ontological category of a lexeme is independent from the semantic role it plays with respect to the other member of the family, it is necessary to separate these two types of information. This is why the structure we propose in Section 6 is articulated on three levels: relational, argumental and ontological.

4 Frame Semantics and FrameNet

Frame Semantics is based on the fundamental assumption that people understand language by means of situations evoked in their mind by words. These representations of real world situations evoked in our mind are called **frames** (Fillmore et al., 1976). For instance, the *Apply_heat* frame describes a common situation involving a COOK, some FOOD and a COOKING_INSTRUMENT and is evoked by lexical units like *bake, blanch, boil, broil, brown, simmer* and *steam*. Lexical units are pairings of words with a meaning. Typically, each sense of a polysemous word belongs to a different frame (Ruppenhofer et al., 2006). For example, the lemma *bake.v* evokes three different frames:

- APPLY_HEAT: Michelle *baked* the potatoes for 45 minutes

- COOKING_CREATION: Michelle *baked* her mother a cake for her birthday

- ABSORB_HEAT: The potatoes have to *bake* for more than 30 minutes

The implementation of Frame Semantics is *FrameNet*, an English lexicon which relates words to their meanings (via the "frames" that they activate) and records the way in which sentences and phrases are structured around them. The main objectives of *FrameNet* are: characterize frames, find the words that evoke those frames, develop a descriptive terminology for each frame and extract sample sentences. Once a frame is defined, it can be used to annotate selected examples from a corpus and to derive valence descriptions for the lexical units involved in the frame itself.

Frames thus represent story fragments, which are evoked by a given set of lexical units (a pairing of a word with a given sense). Each frame is characterised by a certain number of participants involved in it, called **frame elements**. If we take for example the term *avenger* in *FrameNet*, we can see that it evokes the REVENGE frame, whose definition is provided in (3):

(3) An **Avenger** performs a **Punishment** on a **Offender** as a consequence of an earlier action by the Offender, the **Injury**. The **Avenger** inflicting the **Punishment** needs not be the same as the **Injured_party** who suffered the **Injury**, but the **Avenger** does have to share the judgment that the **Offender**'s action was wrong. The judgment that the **Offender** had inflicted an **Injury** is made without regard to the law.

Sentences instantiating this frame:

a. They took REVENGE for the deaths of two loyalist prisoners.
('They' realizes AVENGER and 'for the deaths of two loyalist prisoners' realizes INJURY)

b. Lachlan went out to AVENGE them.
('Lachlan' realizes AVENGER while 'them' realizes INJURED_ PARTY)

c. The next day, the Roman forces took REVENGE on their enemies.
('on their enemies' realizes OFFENDER)

As we can see in (3), the situation is presented by a global definition that shows the core frame elements involved and how they relate with each other. *FrameNet* also provides the non-core frame elements for each frame, which are optional frame elements. For the REVENGE frame, these elements are DEGREE, INSTRUMENT, MANNER, PLACE and PURPOSE. After the global definition of the frame, some example sentences as in (a, b, c) are usually provided in order to show the type of sentences that may instantiate the frame. The second part in the frame representation shown in (4) is composed by partial sentences describing the individual role of each core frame element.

(4) AVENGER: The **Avenger** exacts revenge from the **Offender** for the **Injury**.
    e.g. We want to AVENGE her ('We' realizes AVENGER)

INJURED_PARTY: This frame element identifies the constituent that encodes who or what suffered the **Injury** at the hands of the **Offender**.
e.g. Sam's brothers AVENGED him ('him' realizes INJURED_PARTY)

INJURY: The **Injury** is the injurious action committed by the **Offender** against the **Injured_Party**. This Frame Element needs not always to be realized, although it is conceptually necessary.
e.g. The team sought REVENGE for their 4-1 defeat last night ('for their 4-1 defeat last night' realizes INJURY)

OFFENDER: The **Offender** has committed the earlier **Injury** for which the **Avenger** seeks revenge
e.g. Marie took terrible REVENGE on Trevor ('Trevor' realizes OFFENDER)

PUNISHMENT: The **Avenger** carries out a **Punishment** in order to exact revenge on the **Offender**
e.g. The team took REVENGE with a resounding victory ('with a resounding victory' realizes PUNISHMENT)

Frames also provide information for what concerns the semantic types of the frame elements, even though not all the elements are associated to a semantic category. As far as the REVENGE frame is concerned, the semantic types associated with frame elements (both core and non-core) are provided in (5).

Avenger	Sentient
Instrument	Physical entity
(5) Purpose	State_of_Affairs
Manner	Manner
place	Locative_relation

Last but not least, *FrameNet* also lists all the lexical units that can evoke the frame. For example, for the REVENGE frame, these lexical units are presented in (6):

(6) *avenge.v, avenger.n, get back (at).v, get even.v, payback.n, retaliate.v, retaliation.n, retribution.n, retributive.a, retributory.a, revenge.n, revenge.v, revengeful.a, revenger.n, sanction.n, vengeance.n, vengeful.a, vindictive.a*

5   How could frames be used for morphosemantic description?

As we have seen, *FrameNet* manages to represent a given conceptual situation in an unique object with frames. For what concerns us, our objective is to find a semantic representation for derivational families in a paradigmatic context. One aspect to keep in mind is that *FrameNet* is a resource for English, while *Démonette* is a lexical resource for French.

We can interpret the elements of the derivational family like frame elements in *FrameNet* and put the lexemes of a family in a frame-like structure. In a second moment, we can find other families that fit the same structure and align them, in order highlight regularities in the lexicon and represent a paradigmatic system.

(7) *FrameNet*:

An **Avenger** performs a **Punishment** on a **Offender** as a consequence of an earlier action by the Offender, the **Injury**...

(8) *Démonette*:

Un **laveur lave** quelque chose dans un **lavoir**...
'A washer washes something in a wash house...'

After having created frame-like structures with two or more elements of a derivational family like in (8), we can create an abstract definition by replacing the lexemes with their ontological type and semantic role in square brackets, as in (9):

(9) *Démonette*:

Un **[agent;human][predicate;activity]** quelque chose dans un **[place;artifact]**
'A [agent;human][predicate;activity] something in a [place;artifact]'

We would then have a number of other derivational subfamilies that fit the structure in (9), where the elements of the family would align with the abstract definition like in a paradigmatic system. The next section shows how four derivational families in French could be represented with a frame-like structure.

6   Building a frame-like structure for morphosemantic description

The first family taken as example is the partial family of *laver*, composed by the elements in (10), associated with a morphologically constructed meaning:

	laver	-
	laveur, laveuse	person who washes
	lavoir, laverie	public place where people do the laundry
(10)	lavette	hard sponge used for washing
	lavable	able to be washed
	lavement	procedure / medicinal product for the intestinal washing
	lavage	action or result of the action of washing

As explained in section 3, the description of the derivational family must be structured on three levels of analysis: ontological (which semantic types can be associated to the family elements), relational (how the family elements relate with each other in the sentence) and argumental (which kind of semantic roles are instantiated by the family elements).

6.1 Ontological level

In order to associate the member of the derivational families to a semantic type, a reference ontology needs to be chosen. The basic ontology we used are the **unique beginners for nouns** proposed by Wordnet (Miller, 1995), a large database of English that groups nouns, verbs and adjectives into sets of cognitive synonyms (synsets). In the taxonomy used by Wordnet for nouns, the unique beginners are 25 semantic primes that cover distinct conceptual and lexical domains (Miller et al., 1990). The complete list can be found in (11):

	act, activity	communication	motivation, motive	process
	animal, fauna	event, happening	natural object	quantity, amount
	artifact	feeling, emotion	natural phenomenon	relation
(11)	attribute	food	person, human being	shape
	body	group, grouping	plant, flora	state
	cognition, knowledge	location	possession	substance
	time			

On the ontological level of the representation we propose, each lexeme in a family is associated with a unique beginner, as for the case of *laver* illustrated in (12):

	laver	activity
	lavage	activity
	laveur, **laveuse**	human
(12)	**lavoir**, **laverie**	artifact
	lavable	attribute
	lavette	artifact
	lavement	substance

6.2 Relational level

The information provided on the relational level shows how the family elements relate to each other by means of sentences like those of *FrameNet* including two or more members of a derivational family. Sentences in (13) contain two elements of the family:

(13) a. Un **laveur lave** quelque chose
   b. Une **laveuse lave** quelque chose
   'A washer washes something'

   c. Quelque chose est **lavable** si on peut la **laver**
   'Something is washable if it can be washed'

20

d. On **lave** quelque chose dans une **laverie**
'Something is washed in a laundromat'
e. On **lave** quelque chose dans un **lavoir**
'Something is washed in a wash house'

f. Un **laveur** procède au **lavage** de quelque chose
g. Une **laveuse** procède au **lavage** de quelque chose
'A washer does the washing of something'

h. On réalise un **lavage** quand on **lave** quelque chose
'A washing is realised when we wash something'

i. On pratique un **lavage** sur quelque chose qui est **lavable**
'The washing is done on something that can be washed'

j. Un **lavement lave** l'intestin
'An enema washes the intestine'

k. On realise le **lavage** de quelque chose avec une **lavette**
'We do the washing of something with a dishcloth'

If we take a look at the binary sentences we have constructed, we can see that certain elements will be easier to put together (LAVEUR, LAVEUSE, LAVAGE, LAVETTE), on the other hand it will be almost impossible to combine LAVEMENT with the others, being it a lexeme which refers to a specific medical procedure.

In (14) we present some examples of sentences with three or four elements we can compose:

(14) a. Un **laveur lave** quelque chose dans un **lavoir**
b. Une **laveuse lave** quelque chose dans un **lavoir**
'A washer washes something in a wash house'

c. Quelque chose est **lavable** si un **laveur** peut la **laver**
d. Quelque chose est **lavable** si une **laveuse** peut la **laver**
'Something is washable if a washer can wash it'

e. Un **laveur** fait le **lavage** de quelque chose avec une **lavette**
f. Une **laveuse** fait le **lavage** de quelque chose avec une **lavette**
'A washer does the washing of something with a dishcloth'

g. Un **laveur lave** quelque chose dans un **lavoir** avec une **lavette**
h. Une **laveuse lave** quelque chose dans un **lavoir** avec une **lavette**
'A washer washes something in a wash house with a dishcloth'

6.3 Argumental level

The representation also provides the semantic role of each element with respect to the other members of the family. The argumental level associates the element of the family with their role in the argumental structure, which has been deducted from the category of relation where they are inscribed.

laver	predicate
lavage	pred. with support verb (pratiquer/faire)
(15) **laveur/laveuse**	agent
lavoir/laverie	place
lavette	instrument
lavable	modifier

Relating *lavement*, it results to be difficult to place it in the same structure with the other elements of the subfamily because it poses a polisemy problem: the sense of *laver* in relation with *lavement* is not the same as *laver* when it is considered in relation with the other elements of the family. This is why it needs to be considered separately, since the only relation where it is involved is the one with *laver*:

(16)

laver	predicate
lavement	pred. with support verb (administrer/faire)

6.4 Catching paradigmatic generalizations

Our frame-like representation is also fit for the representation of the paradigmatic organization of the derivational lexicon. We tested the structure we built for *laver* on three other subfamilies: *observer* 'to observe', *imprimer* 'to print' and *nettoyer* 'to clean'. The three families we chose are articulated around verbs concerning human activities, like *laver*. In table 2 we present the other three subfamilies:

observer 'to observe'	observateur , 'observer(m.)'	observatrice 'observer' (f.)	observation , 'observation'	observable 'observable'	observatoire 'observatory'
imprimer 'to print'	imprimeur 'printer'(m.)	imprimeuse 'printer' (f.)	impression 'printing'	imprimable 'printable'	imprimerie 'copy shop'
nettoyer 'to clean'	nettoyeur 'cleaner' (m.)	nettoyeuse 'cleaner' (f.)	nettoyage 'cleaning'	nettoyable 'cleanable'	–

Table 2: (sub) families of *observer*, *imprimer* and *nettoyer*

The alignment in Table 2 can be extended to the relational level as in from tables 3 to table 10. The bottom line in each table provides an abstract definition where the lexemes are abstracted by the combination of the semantic role and the ontological type in square brackets.

Un **laveur**	**lave**	quelque chose
Un **nettoyeur**	**nettoie**	quelque chose
Un **observateur**	**observe**	quelque chose
Un **imprimeur**	**imprime**	quelque chose
Un **[agent; human m.]**	**[predicate; activity]**	quelque chose

Table 3: masculine human agent and activity

Une **laveuse**	**lave**	quelque chose
Une **nettoyeuse**	**nettoie**	quelque chose
Une **observatrice**	**observe**	quelque chose
Une **imprimeuse**	**imprime**	quelque chose
Une **[agent; human f.]**	**[predicate; activity]**	quelque chose

Table 4: feminine human agent and activity

On **lave**	quelque chose dans une	**laverie** **lavoir**
On **imprime**	quelque chose dans une	**imprimerie**
On **observe**	quelque chose dans un	**observatoire**
On **nettoye**	quelque chose dans un	**?**
On **[predicate;activity]**	quelque chose dans un/une	**[place; artifact]**

Table 5: activity and artifact

Quelque chose	est **lavable**	si on peut	la **laver**
Quelque chose	est **imprimable**	si on peut	l'**imprimer**
Quelque chose	est **observable**	si on peut	l'**observer**
Quelque chose	est **nettoyable**	si on peut	la **nettoyer**
Quelque chose	est **[modifier; attribute]**	si on peut	la/le/l' **[predicate; activity]**

Table 6: attribute and activity

The alignments in tables 3, 4 and 6 are complete, while in table 5 the family of *nettoyer* lacks a member denoting the place where the cleaning takes place. Sentences containing three elements could be aligned similarly:

Un **imprimeur**	**imprime**	quelque chose	dans une **imprimerie**
Un **observateur**	**observe**	quelque chose	dans un **observatoire**
Un **laveur**	**lave**	quelque chose	dans une **laverie**
Un **nettoyeur**	**nettoie**	quelque chose	dans une **?**
Un **[agent; human m.]**	**[predicate; activity]**	quelque chose	dans un/une **[place; artifact]**

Table 7: human masculine agent, activity and artifact

Une **imprimeuse**	**imprime**	quelque chose	dans une **imprimerie**
Une **observatrice**	**observe**	quelque chose	dans un **observatoire**
Une **laveuse**	**lave**	quelque chose	dans une **laverie**
Une **nettoyeuse**	**nettoie**	quelque chose	dans une **?**
Une **[agent; human f.]**	**[predicate; activity]**	quelque chose	dans un/une **[place; artifact]**

Table 8: human feminine agent, activity and artifact

Quelque chose est	**imprimable**	si un **imprimeur**	peut l'**imprimer**
Quelque chose est	**observable**	si un **observateur**	peut l'**observer**
Quelque chose est	**lavable**	si un **laveur**	peut la **laver**
Quelque chose est	**nettoyable**	si un **nettoyeur**	peut la **nettoyer**
Quelque chose	est **[modifier; potentiality]**	si un **[agent; human m.]**	peut la/l' **[predicate; activity]**

Table 9: Modifier, human masculine agent and activity

Quelque chose est	**imprimable**	si une **imprimeuse**	peut l'**imprimer**
Quelque chose est	**observable**	si une **observatrice**	peut l'**observer**
Quelque chose est	**lavable**	si une **laveuse**	peut la **laver**
Quelque chose est	**nettoyable**	si une **nettoyeuse**	peut la **nettoyer**
Quelque chose	est **[modifier; potentiality]**	si une **[agent; human f.]**	peut la/l' **[predicate; activity]**

Table 10: Modifier, human feminine agent and activity

As we can see, the alignment in Table 7 and Table 8 is partial due to the absence in the family of *nettoyer* of a lexeme denoting the location where the action takes place. In Table 9 and Table 10, on the other hand, the alignment works for the four families.

7   Conclusions

We showed that semantic frames used by *FrameNet* can be easily adapted to represent a derivational family and can also represent alignments of families and derivational paradigms. The next step consists in developing a program capable of building frame-like representations from lexical, lexicographic and distributional data.

Acknowledgements

References

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley Framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

Bonami, O. and Strnadová, J. (2018). Paradigm structure and predictability in derivational morphology. *Morphology*, pages 1–31.

Fillmore, C. J. et al. (1976). Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280, pages 20–32.

Fillmore, C. J. et al. (2006). Frame semantics. *Cognitive linguistics: Basic readings*, pages 373–400.

Haspelmath, M. and Sims, A. (2013). *Understanding morphology*, pages 2–3. Routledge.

Hathout, N. (2009). Acquisition of morphological families and derivational series from a machine readable dictionary.

Hathout, N. and Namer, F. (2014). Démonette, a french derivational morpho-semantic network. *LiLT (Linguistic Issues in Language Technology)*, 11.

Hathout, N. and Namer, F. (2016). Giving lexical resources a second life: Démonette, a multi-sourced morpho-semantic network for french. In *Language Ressources and Evaluation Conference*.

Hathout, N., Namer, F., and Lignon, S. (2017). Adding morpho-phonological features to a french morphosemantic resource: the demonette derivational database. In *Workshop on Resources and Tools for Derivational Morphology (DeriMo)*.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4):235–244.

Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C. R., and Scheffczyk, J. (2006). Framenet ii: Extended theory and practice.