Frame-like structure for morphosemantic description

Daniele Sanacore, Nabil Hathout, Fiammetta Namer

FRAME-LIKE STRUCTURES FOR MORPHOSEMANTIC DESCRIPTION

Daniele SANACORE
Université de Toulouse

Nabil HATHOUT
Université de Toulouse

Fiammetta NAMER
Université de Lorraine

RÉSUMÉ

Cet article présente un nouveau formalisme pour la représentation des relations morphosémantiques dans le lexique construit du français, dans le cadre de la morphologie paradigmatique. Ce travail alimentera Démonette, une base de données morphologiques de grande taille. Nous proposons dans cette étude d'utiliser des structures inspirées des frames de la Sémantique des Frames pour décrire globalement l'ensemble des relations de sens qui s'établissent au sein des familles dérivationnelles et pour mettre en évidence la nature paradigmatique du lexique construit.

ABSTRACT

In the framework of paradigmatic morphology, this article presents a new formalism for the representation of morphosemantic relations in the French derivational lexicon. This work will feed Démonette, a large derivational database describing word formation in French. Our approach is inspired by Frame Semantics: we show that morphosemantic frames can be used to describe the semantic relations that hold in a derivational family. Moreover, these frames can be aligned in order to form morphosemantic paradigms and highlight the paradigmatic nature of the derivational lexicon.

1. INTRODUCTION

In this article we propose a new formalism for the representation of semantic relations in derivational families (i.e. sets of derivationally related words) in the French constructed lexicon. We call this formalism "Morphosemantic Frames" (MFs) because it is inspired by the frames of Frame Semantics (Fillmore, 1976; Fillmore & Baker, 2001).

In this work we focus on the morphosemantic analysis of derivational families and their description. To our knowledge, derivational morphology lacks a formalism able to describe the morphosemantic relations between the lexemes of derivational families and that account for their regularity in the lexicon. Morphosemantic Frames aim at filling this gap. Moreover, this formalism could be used for other languages in order to develop a common framework for the morphosemantic representation of derivational relations in their constructed lexicon.

MFs focus on the morphosemantic relations between lexemes in the derivational families. The framework is designed to represent the structural regularity of the lexicon, by showing that the same groups of semantic relations connect the members of several families. MFs are precise, economic and have a high predictive power. They account for all the semantic relations between the lexemes contained in derivational families; they are easy to read and at the same time are applicable to large datasets; MFs allow for the prediction of missing lexemes and for the reconstruction of the derivational families.

Following (Hathout & Namer 2014), we consider that the meaning of a given lexeme is described by the contribution of each of the derivational relations where it is involved. For instance, the examples in (1), (2) and (3) show how several derivational relations between a verb and its derivatives determine its argument structure.

(1) *ronfler*.V 'snore'- *ronfleur*.N 'person who snores' *chanter*.V 'sing' - *chanteur*.N 'singer'

In (1), we deduce from the relation between the verb and its derived noun in *-eur* that it has an agent argument and the derived noun denotes an instrument or a person that performs the action (*chanteur* and *ronfleur*).

(2) *périr*.V 'perish', *périssable*.A 'perishable'

nager.V 'swim', *nageable*.A 'swimmable'

modifier.V 'modify', *modifiable*.A 'modifiable'

In (2), we can deduce from the relation between the verbs and their adjectives in *-able* that these verbs have a patient (in the case of *perishable* or *modifiable*) or a locative modifier (in the case of *nageable*). We therefore

see how different derivational relations of a verb contribute to the identification of the content of its argument structure.

(3) *pêcher*.V 'fish', *pêchable*.A 'fishable'

In some cases, as with *pêchable* in (3) there is more than one possible interpretation*: pêchable* can be associated with the patient of the process, with the place where the process occurs, with the time period where the action is performed or with the instruments that are used (Hathout et al., 2004), which are all internal arguments.

(4) *laver*.V 'wash', *laveur*.N 'washer', *lavable*.A 'washable'

présenter.V 'present', *présentateur*.N 'presenter', *présentable*.A 'presentable'

The configuration in (4) presents both the derived noun in -*eur* and the derived adjective in -*able*. The access to the entire family in (4) allows us to get more insight about the argument structure of the base verb than single word pairs as in (1), (2) and (3). We therefore need to provide descriptions of the entire derivational families that go beyond individual base-derivative couples.

Morphosemantic analysis also involves ontological properties. Let us consider two derivational families: the one of *banane* 'banana' in (5) and the one of *balai* 'broom' in (6). Even though these two families have two nouns as their roots, they are rather different in the concepts that they express.

(5) *banane*.N 'banana', *bananier*.N 'banana tree', *bananeraie*.N 'banana plantation',

(6) *balai*.N 'broom', *balayeur*.N 'male sweeper', *balayer*.V 'sweep', *balayage*.N 'sweeping'

The difference stems from the ontological nature of their root noun: *banane* denotes a fruit and yields nouns like *bananier* and *bananeraie* which indicate respectively the plant and the plantation that produces that fruit. The noun *balai* denotes an artifact and its family includes the derived agent in -*eur*, which stands for a person that uses it. The verb *balayer* and the noun *balayage* denote the action of using the instrument. These examples show that the ontological categories determine the architecture of the derivational families. This is confirmed by other families of words derived from fruits in (7) and (8) and from artifacts in (9) and (10):

(7) *cerise.*N 'cherry', *cerisier.*N 'cherry tree', *ceriseraie.*N 'cherry plantation'

(8) *amande.*N 'almond', *amandier.*N 'almond tree', *amanderaie.*N 'almond plantation'

(9) *chronomètre.*N ‘chronometer’*, chronométrer.*V ‘chronometre’, *chronométreur.*N ‘person who chronometers’ , *chronométrage.*N ‘time keeping’

(10) *brosse.*N ‘brush’, *brosser.*V ‘brush’, *brosseur.*N ‘person who brushes’, *brossage.*N ‘brushing’

The families in (7) and (8) contain the same concepts and relations between concepts as the family of *banane*: they both include the noun for the tree (*cerisier* and *amandier*) and the noun of the plantation that produces the fruit (*ceriseraie* and *amanderaie*). Families (9) and (10), contain the same concepts as the family of *balai* in (6): the human agent (*chronométreur* and *brosseur*) and the action in verbal (*chronométrer* and *brosser*) and nominal forms (*chronométrage* and *brossage*).

These examples show that the ontological categories of the lexemes play an important role in the analysis and description of derivational families and that they must be represented in the Morphosemantic Frames.

Another difference between the families in (5), (7) and (8), and in (6), (9) and (10) is that the latter are event-related (their concepts denote the participants of some action) and that the former are not, as outlined in (Fradin, in press). The concepts of event-related families are therefore also relevant for the syntax.

The remaining of the paper is structured as follows: Section 2 outlines the theoretical background of MFs: paradigmatic morphology. Section 3 introduces Frame Semantics, which has inspired our formalism. Section 4 shows how the constitutive elements of Frame Semantics are adapted to MFs and used for morphosemantic descriptions. Section 5 describes the structure of MFs and shows how we represent the families and their structures. Section 6 discusses the implementation of Morphosemantic Frames in the derivational lexicon *Démonette*.

2. PARADIGMATIC MORPHOLOGY

As we just saw, derivational families are a fundamental object in our analysis and we need a theoretical framework that provides tools to describe and handle them. Paradigmatic derivational morphology responds precisely to this need. Paradigmatic derivational morphology accounts for the numerous regularities that are present in the lexicon (Van Marle, 1984; Stump 1991; Bauer, 1997; Booij, 2008 *inter alia*) and organizes families into paradigms.

A derivational family is a set of derivationally related lexemes (Hathout, 2011) as the ones in (11). In the following, the term ‘family’ is also used for subsets of derivational families (Bonami & Strnadova, 2019).

(11) *présenter*.v ‘present’, *présentation*.N ‘presentation’, *présentateur*.N ‘presentermasc’, *présentable*.A ‘presentable’

In paradigmatic morphology, a paradigm is a set of derivational families structured around the same oppositions of content (Bonami & Strnadova, 2019). MFs being morphosemantic descriptions, we will only be concerned with morphosemantic paradigms, which means that we are interested only in the concepts expressed by the lexemes present in derivational families and the semantic relations between these concepts. For example, Table 1 presents three families that we can include in the same morphosemantic paradigm since the lexemes that they contain express the same concepts, regardless of the fact that these concepts are realized by different formal means (conversion for *vol*, suffixation in *-ion* for *présentation* and in *-age* for *lavage*). These three families are all made up of a verb expressing an event, an agent noun and a noun denoting the action.

event verb	**agent noun**	**action noun**
présenter.V ‘present’	*présentateur*.N ‘presenter$_{masc}$’	*présentation*.N ‘presentation’
voler.V ‘steal’	*voleur*.N ‘thief$_{masc}$’	*vol*.N ‘theft’
laver.V ‘wash’	*laveur*.N ‘washer$_{masc}$’	*lavage*.N ‘washing’

Table 1. – Example of morphosemantic paradigm

The families in (5), (6), (7), (8), (9) and (10) can be grouped in two paradigms, each one being defined by the sets of semantic relations shared by (5, 7, 8) and (6, 9, 10).

Figure 1 represents the paradigm containing the families of *banane*, *cerise* and *amande*. The three families contain the same concepts and the same relations between these concepts. They are thus aligned under the same paradigm. Moreover, other families rooted in fruit nouns can be aligned under the same morphosemantic structure. Alignment is a key feature for paradigms because it gives them their predictive power, one of the features that we need for our formalism. Two pairs of morphologically related words (w1, w2) and (w3, w4) are aligned when the two pairs are linked by the same content relation (Bonami & Strnadova, 2019). When another noun is analysed as a fruit name (e.g. *abricot* ‘apricot’), the predictive power of the paradigm allows us to deduce that its family also contains a name for the plant (*abricotier* ‘apricot tree’) and a name for the plantation that produces this fruit (*abricoteraie* ‘apricot plantation’) .

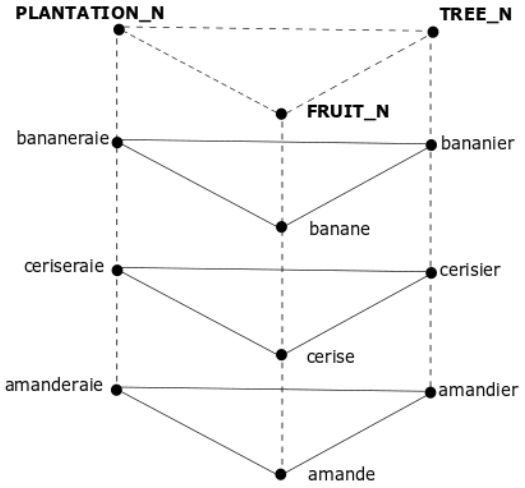

Figure 1. – Fruit paradigm

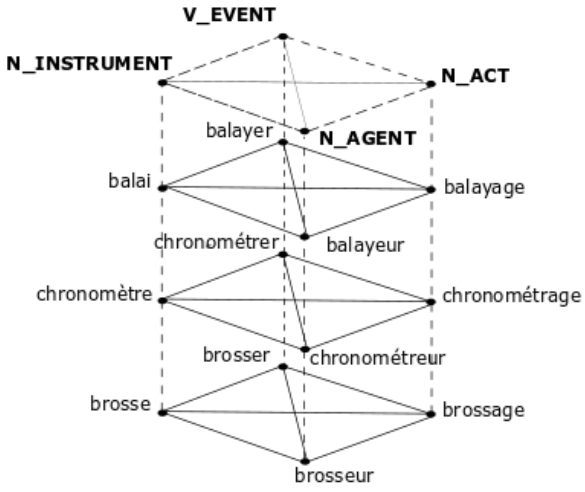

Figure 2. – Artifact paradigm

The paradigm in Figure 2 represents the alignment of the families in (6), (9) and (10) with an instrument noun, the agent that makes use of that instrument and the action in nominal and verbal forms. As outlined in the

introduction, the concepts expressed in this paradigm are relevant for the syntax-semantics interface. This is not the case in the *Fruit Paradigm*.

In both examples we have an abstract structure (the morphosemantic paradigm) represented at the top of each figure, and its realizations (the families) represented by graphs formed by continuous lines. In the derivational families, the lines represent morphological relations between the lexemes. As it will be seen in section 4, Morphosemantic Frames describe morphosemantic paradigms like those in Figures 1 and 2.

3. FRAME SEMANTICS

This section presents Frame Semantics and the concepts borrowed and adapted for morphosemantic description. The theoretical framework of Frame Semantics is based on the assumption that words represent categories of experience and evoke in the mind of the listener representations of real-world situations, called frames. The best-known implementation of Frame Semantics is *FrameNet* (Baker et al., 1998; Ruppenhofer et al., 2006), a resource designed for the semantic annotation of English texts. *FrameNet* has inspired the creation of equivalent resources in several other languages (Lenci et al. 2010; Candito et al., 2014 *inter alia*).

In *FrameNet* a frame is characterized by a number of frame elements (FEs), which correspond to situational roles that are characteristic of the situation described in the frame. For example, the ARREST frame has *Authorities*, *Charges* and *Suspect* as frame elements. The frame is first described by a sentence (the frame description) that globally defines the represented cognitive situation and how the frame elements are involved in it (12). It also contains partial descriptions that present the cognitive situation from the point of view of each frame element (13). In addition, some FEs are associated with ontological categories (e.g. *Authorities* is given the *Sentient* label).

(12) ARREST: **Authorities** charge a **Suspect**, who is under suspicion of having committed a crime (the **Charges**), and take him/her into custody

(13) AUTHORITIES (Sentient): The **Authorities** charge the **Suspect** with committing a crime and take him/her into custody.

CHARGES: **Charges** identifies a category within the legal system; it is the crime with which the **Suspect** is charged

SUSPECT: The **Suspect** is taken into custody, under suspicion of having committed a crime.

Frames are also characterized by lexical units (LUs), that is lemmas that evoke the cognitive situation described by the frame. The LUs of the ARREST frame are given in (14).

(14) apprehend.V, apprehension.N, arrest.N, arrest.V, book.V, bust.N, bust.V, collar.V, cop.V, nab.V, summons.V

They also contain examples of sentences that realize them in texts. Several sentences like those in (15), (16) and (17) may evoke the same cognitive structure (the same frame) and illustrate the use of different LUs. Corpus sentences are thus the concrete realization of the frame.

(15) The police ARRESTED Harry on charges of manslaughter.

(16) The seven were BOOKED on marijuana possession charges.

(17) Almost four million people in England and Wales were SUMMONSED in the first half of this year for failing to pay.

To summarize, in *FrameNet*, frames are composed of frame elements, frame descriptions, lexical units and are realized by corpus sentences. In the semantic frames, frame elements and frame descriptions represent abstract structures instanciated by the lexical units and the corpus sentences. This abstract/concrete distinction is also key for Morphosemantic Frames and derivation families, as showed in section 4.

4. MORPHOSEMANTICS DESCRIPTION WITH FRAMES

In *FrameNet* different sentences containing different lexical units may be structured in the same way when they evoke the same cognitive structure (the same frame). An analogous configuration exists in morphosemantics: different families are structured on the same meaning oppositions and instanciate (evoke) the same morphosemantic paradigm, as shown in Figures 1 and 2. Morphosemantic Frames represent abstract paradigms instanciated by derivational families (aligned in concrete paradigms).

MFs include 'Morphosemantic Glosses' (MGs) which describe how the concepts of the morphosemantic paradigm are related to one another. MGs are sentences similar to the frame descriptions in *FrameNet*. For instance, the abstract paradigm in Figure 1 can be described with the Morphosemantic Glosses in (18).

(18) *Une **Plantation** produit des **Fruits***

'A Plantation produces some Fruits'

*Une **Plantation** ne comporte que des **Plantes** de la même espèce*

'A Plantation only contains Plants of the same species'

*Un **Fruit** pousse sur une **Plante***

'A Fruit grows on a Plant'

The frame elements in the Morphosemantic Glosses can be instanciated with the lexemes of the families of the paradigm. (19) shows how these lexemes of the family of *banane* fit in the structure of the Morphosemantic Glosses.

(19) *Une **bananeraie** produit des **bananes***

'A banana plantation produces bananas'

*Une **bananeraie** ne comporte que de **bananiers***

'A banana plantation only contains banana trees'

*Une **banane** pousse sur un **bananier***

'A banana grows on a banana tree'

We can do the same for all the other families aligned under the morphosemantic paradigm in Figure 1 (e.g. *cerise* or *amande*) and to the artifact paradigm in Figure 2. Its morphosemantic glosses and concrete sentences are presented in Table 2.

Morphosemantic Glosses	**Concrete sentences**
Un Agent utilise un Instrument 'An Agent uses an Instrument'	*Un balayeur utilise un balai* 'A sweeper uses a broom'
Un Agent réalise l'acte de Événement 'An Agent realizes the action of Event'	*Un balayeur réalise l'acte de balayer* 'A sweeper realizes the act of sweep'
Un Agent réalise l'Action 'An Agent realizes the Action'	*Un balayeur realise le balayage* 'A sweeper realizes the sweeping'
Un Instrument est utilisé pour réaliser une Action 'An Instrument is used to realize the Action'	*Un balai est utilisé pour réaliser le balayage* 'A broom is used to realize the sweeping'
Un Instrument est utilisé pour Événement 'An Instrument is used for Event'	*Un balai est utilisé pour balayer* 'A broom is used for sweep'
L'Action est l'acte de Événement 'Action is the act of Event'	*Le balayage est l'acte de balayer* 'Sweeping is the act of sweep'

Table 2. – MGs and concrete sentences for the Artifact paradigm

In order for these sentences to sound natural and easy to read, we limit the number of frame elements in the MGs to two or three (this is one

difference with the semantic frames). Table 3 summarizes the correspondences between frames in *FrameNet* and MFs. The Morphosemantic glosses used for the description of the morphosemantic abstract paradigms correspond to the frame descriptions in semantic frames. In these glosses, the concepts expressed in the paradigm are related and defined simultaneously like frame elements in frame descriptions. Finally, the derivational families correspond to the lexical units and the concrete interdefining glosses that relate the members of a derivational family correspond to the corpus sentences. They instanciate the morphosemantic glosses just as corpus sentences realize the semantic descriptions.

Semantic Frames	**Morphosemantic frames**
frame description	morphosemantic glosses
frame elements	concepts forming of the paradigm
lexical unit	derivational families
corpus sentences	concrete glosses

Table 3. – Correspondences between frames and MFs

5. THE STRUCTURE OF THE MORPHOSEMANTIC FRAMES

The elements introduced in section 4 make up the Morphosemantic Frames. What do they look like? Morphosemantic frames are structured in three layers: the first contains the Morphosemantic Glosses and a label identifying the corresponding semantic relation, as in Table 4. The objective of this label is defining the semantic relation between the elements in the semantic gloss. The definition of a label set for this purpose is in progress and the proposed labels are temporary. The second layer (Table 6) describes the association between the frame elements and their ontological labels. The third (Table 7) lists the families that realize the morphosemantic paradigm described in the MF.

Morphosemantic glosses+Semantic Relation label	
Une **Plantation** produit des **Fruits**	**Producer - Product**
Une P**lantation** est plantée de **Plantes** de la même espèce	**Group - Entity**
Un **Fruit** pousse sur une **Plante**	**Entity - Plant**

Table 4. – Morphosemantic glosses and semantic labels for the Fruit Paradigm

The second layer associates each concept with an ontological category (Table 6). In this article, we use the categories of the *FrSemCor* project (Barque et al., 2020) for the nouns. A sample is presented in Table 5. These categories are a modified version of the *Unique Beginners for Nouns* used by

Wordnet (Miller, 1998) and present different levels of granularity. They allow a high precision as in the cases of *fruit* and *tree* but also include more generic categories. However, this ontology has not been created for morphosemantics. This is why we intend to adapt it to the morphosemantics of French in the next phases of the research. For example, different categories of people can be expressed morphologically, like the specialists of a given domain (*économiste* 'economist', *philologue* 'philologist'), the people that adhere to a given doctrine (*Marxiste* 'Marxist', *fasciste* 'Fascist') or the shopkeepers (*poissonier* 'fishmonger', *fleuriste* 'florist'). These words are tagged as *Person* with the current classification, which cannot be useful to predict the remainder of their families, e.g. *poissonnerie* in the family of *poissonnier*.

Entity		
	Living & animate_entity	
		Animal (e.g. *fourmilier* 'ant-bear')
		Person (e.g. *défenseur* 'defender')
	Non_Animate_entity	
		Food (e.g. *orangeade* 'orangeade')
		Substance (e.g. *chlorure* 'chlorure')
		Artifact (e.g. *aspirateur* 'vacuum cleaner')
		Plant (e.g. *bananier* 'banana tree')

Table 5. –Sample of the noun ontology in *FrSemCor*

POS+Ontology	
Fruit.N	Fruit
Plante.N	Plant
Plantation.N	Group x Plant

Table 6. – Frame Elements and ontological labels for the Fruit paradigm

In the *Fruit Paradigm*, the label of *Plantation* is *Plant* preceded by the modifier *Group*. The same modifier can be applied for all collective nouns: people, (e.g. *foule* 'crowd'), animals (e.g. *colonie* 'colony'), artifacts (e.g. *armement* 'weapons'), etc. The verb and adjective ontology have not been worked out yet. For the moment, verbs may have two labels (*stative situation* or *dynamic situation*) and all adjectives are tagged as *modifier*. Table 7 presents the families that realize the *Fruit Paradigm* of Figure 1.

Families		
Fruit.N	**Plante.N**	**Plantation.N**

banane	bananier	bananeraie
cerise	cerisier	ceriseraie
amande	amandier	amanderaie

Table 7. – Families that realize the MF of the Fruit paradigm

We present in Tables 8, 9 and 10 the three layers for the morphosemantic frame that represents the *Artifact Paradigm* in Figure 2. In Table 8, the morphosemantic glosses correspond to the content of Table 2, and semantic relations in the paradigm are labelled with general categories like *Agent - Event*, *Instrument - Event*, *Agent - Instrument* and *Synonymy*.

Morphosemantic glosses+Semantic Relation label	
Un **Agent** utilise un **Instrument**	**Agent - Instrument**
Un **Agent** réalise l'acte de **Événement**	**Agent - Event**
Un **Agent** realise l'**Action**	**Agent - Event**
Un **Instrument** est utilisé pour réaliser une **Action**	**Instrument - Event**
Un **Instrument** est utilisé pour **Événement**	**Instrument - Event**
L'**Action** est l'acte de **Événement**	**Synonymy**

Table 8. – Morphosemantic glosses and semantic labels for the Artifact Paradigm

In Table 9, ontological classes assigned to the concepts of the *Artifact paradigm* correspond to the more specific values in the UB hierarchy: instruments are *Artifacts;* agents are human beings (labelled with the category *Person*) and events belong to the class of *Dynamic situation.*

POS+Ontology	
Instrument.N	Artifact
Agent.N	Person
Événement.V	Dynamic situation
Action.N	Act

Table 9. – Frame Elements and ontological labels for Artifact paradigm

Table 10 contains the families of Figure 2; they realize the *Artifact Paradigm*.

Families			
Instrument.N	**Evenément.V**	**Agent.N**	**Action.N**
balai	balayer	balayeur	balayage
chronomètre	chronométrer	chronométreur	chronométrage
brosse	brosser	brosseur	brossage

Table 10. – Families that realize the MF for Artifact paradigm

6. THE DESCRIPTION OF MORPHOSEMANTIC FRAMES IN DEMONETTE

The *Démonette* database is a derivational resource that will be fed by this work. In this section, we outline its constitutive elements and show how Morphosemantic Frames will be implemented in *Démonette*. *Démonette* (Hathout & Namer, 2014, 2016; Namer et al, 2019) is a large-scale derivational database for French. It implements the theoretical model *ParaDis* (Namer & Hathout, 2020), where lexemes are grouped into families, which in turn are organized into paradigms. The design of *Démonette* is based on a cumulative conception of meaning, as presented in the introduction: the meaning of a derived word is the combination of the semantic properties of all the derivational relations where the word is involved. *Démonette* has four objectives (Hathout & Namer, 2014): (i) connect the members of a derivational family by direct and indirect relations; (ii) label each relation semantically; (iii) provide the words in the database with phonological, morphological and semantic information; (iv) provide similar definitions to the words that are part of the same paradigms. *Démonette* is fed by several existing resources of different nature (Namer et al., 2019). In the *Démonette* database, an entry corresponds to a binary relation between two derivationally related lexemes lex1 and lex2 (e.g. *laveur* 'washer' and *laver* 'wash'). Contrarily to the relations in Morphosemantic Frames, in *Démonette* relations are oriented (e.g. *laver* → *laveur* and *laveur* →*laver*). The table that realizes the database includes the ontological types of lex1 and lex2 and the semantic relation between lex1 and lex2 as illustrated in Table 11.

Lex1-Lex2	**Onto1**	**Onto2**	**Sem. Rel.**	**Abstract Definition**
bananeraie-bananier	Group x Plant	Plant	collection	*Un Lex2 est un ensemble de Lex1* 'A Lex2 is a collection of Lex1'
balayer-balayage	dyn-situation	act	synonymy	*Lex1 c'est faire Lex2* 'Lex1 is doing Lex2'

Table 11. – Semantic description in Démonette

On the other hand, the derivational database contains elements that can be used to reconstitute Morphosemantic Frames. *Démonette* provides a numerical identifier of the family of the lexemes, which means that the families can be recovered from the database. *Démonette* uses the same

ontological labels as the MFs. *Démonette* also describes the semantic relations that hold between the pairs of lexemes, as shown in the fourth column in Table 11 (however, the label is single and not double) and gives an abstract definition in the last column. Finally, the semantic description is complemented by two glosses, a concrete and an abstract paraphrases which inter-define the two lexemes (Namer et. al, 2019) as shown in Table 12.

(Lex1-Lex2)	Concrete paraphrase	Abstract paraphrase
bananeraie-bananier	Une bananeraie ne comporte que des bananiers	Une *Group x Plant.N* ne comporte que des *Plant.N*
balayer-balayage	Le balayage est l'acte de balayer	*act.N* est l'acte de *sit_dyn.V*

Table 12. – Concrete and abstract paraphrases in *Démonette*

These sentences have the same structure as the Morphosemantic Glosses. Morphosemantic Glosses could thus be added to the database in order to fill in the semantic relations that are included in the same paradigm.

We can therefore fill in *Démonette* with the Morphosemantic Frames in a quite straightforward way. On the other hand, some information already in the database can be used to create the MFs: the binary relations between lexemes, the family of each binary relation in the database, the ontological labels and the concrete and abstract paraphrases.

7. CONCLUSION

In this article we presented our formalism for the representation of morphosemantic lexicon: the Morphosemantic Frames. MFs provide a semantic description of the derivational relations and an ontological categorization of the lexemes. We showed how derivational families instanciate the semantic paradigms described in the Morphosemantic Frames and highlighted the parallelism between the MFs and the semantic frames of *FrameNet*.

The adaptation of the constitutive elements of frame semantics to our task has been proven to be solid and capable of bringing out structural regularities in the lexicon. Moreover, we showed the main advantages of MFs with some examples: their readability, the fact that they can be applied on large-scale data and their predictive power. We described the type of morphosemantic information that we need to convey in our MFs and how we express it (abstract glosses, ontological labels and relation labels).

In the near future, our objective is to develop a large enough number of frames like those proposed in this article and describe on a large scale the data contained in *Démonette*. On the longer term, we aim at something like a

FrameNet for morphosemantics (in a prototypical form) and its implementation in *Démonette* in a first moment. Fine-grained description would also help us deal with lexical gaps, a problem that we will need to be addressed in the frames.

Another important question we intend to address concerns the ontological categories needed for morphosemantics and the labelling of the semantic relations. The ontology must provide categories adapted for verbs and adjectives. Our approach will be bottom-up, starting from the direct observation of data.

RÉFÉRENCES

BAKER, C. F., FILLMORE, C. J., & LOWE, J. B. (1998). The Berkeley Framenet project. *In Proceedings of the 17th international conference on Computational linguistics-Volume 1 86-90. Association for Computational Linguistics*.

BARQUE L., HAAS P., HUYGHE, R., TRIBOUT, D., CANDITO, M., CRABBÉ, B., & SEGONNE, V. (2020). FrSemCor: Annotating a French corpus with supersenses. In *LREC-2020*.

BAUER, L. (1997). Derivational paradigms. In *Yearbook of Morphology* 243-256. Springer, Dordrecht.

BONAMI O., & STRNADOVÁ, J. (2019). Paradigm structure and predictability in derivational morphology.*Morphology*,29(2), 167-197.

BOOIJ, G. (2008). Paradigmatic morphology. *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, 29-38. John Benjamins Publishing.

CANDITO M., AMSILI P., BARQUE L., BENAMARA F., DE CHALENDAR G., DJEMAA M. & SAGOT, B. (2014, May). Developing a french framenet: Methodology and first results. *In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*

FILLMORE, C. J. (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 (Vol. 280, No. 1)* 20-32.

FILLMORE C. J. & BAKER, C. F. (2001). Frame semantics for text understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL (Vol. 6)*.

FRADIN B (to appear). *On the variety of derivational paradigms. in Paradigmatic relations in derivational morphology*. Bagasheva, Alexandra, Fernández-Domínguez, Jesús, Lara-Clares, Cristina (eds). Leiden: Koninklijke Brill NV, In press(halshs-01955558).

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. *LiLT (Linguistic Issues in Language Technology), 11*.

HATHOUT N., PLÉNAT, M., & TANGUY, L. (2004). *Enquête sur les dérivés en-able. Cahiers de grammaire* 28, 49-90.

HATHOUT, N., & NAMER, F. (2016). Giving lexical resources a second life: Démonette, a multi-sourced morpho-semantic network for French. *Language Ressources and Evaluation Conference*, May 2016, Portoroz, Slovenia.

HATHOUT, N. (2011). Morphonette: a paradigm-based morphological network, in *Lingue e linguaggio, Rivista semestrale* 2/2011, pp. 245-264.

LENCI, A., JOHNSON, M., & LAPESA, G. (2010). Building an Italian FrameNet through Semi-automatic Corpus Analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation* (LREC '10).

MILLER, G. A. (1998). *WordNet: An electronic lexical database*. MIT press.

NAMER F., BARQUE, L., BONAMI O., HAAS P., HATHOUT, N., & TRIBOUT, D. (2019). Demonette2 - Une base de données dérivationnelles du français à grande échelle: premiers résultats. *Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN) - PFIA 2019 - Volume II* : Articles courts, Jul 2019, Toulouse, France.

NAMER F., HATHOUT N. (2020). ParaDis and Démonette-From Theory to Resources for Derivational Paradigms. *The Prague Bulletin of Mathematical Linguistics* 114, 5-33.

RUPPENHOFER J., ELLSWORTH M., SCHWARZER-PETRUCK M., JOHNSON C. R., & SCHEFFCZYK J. (2006). *FrameNet II: Extended theory and practice*. (revised ed. November 1, 2016). Berkeley, CA: International Computer Science Institute.

STUMP G.T. (1991). A paradigm-based theory of morphosemantic mismatches. *Language*, Linguistic Society of America 675-725.

VAN MARLE J. (1984). *On the paradigmatic dimension of morphological creativity*. Berlin, Boston: De Gruyter.