ParaDis and Démonette – From Theory to Resources for Derivational Paradigms

Fiammetta Namer, Nabil Hathout

▶ **To cite this version:**

Fiammetta Namer, Nabil Hathout. ParaDis and Démonette – From Theory to Resources for Derivational Paradigms. The Prague Bulletin of Mathematical Linguistics, 2020, 114 (1), pp.5-34. 10.14712/00326585.001 . halshs-02707601

HAL Id: halshs-02707601
https://shs.hal.science/halshs-02707601

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PBML

The Prague Bulletin of Mathematical Linguistics
NUMBER 114 APRIL 2020 5–33

ParaDis and Démonette. From Theory to Resources for Derivational Paradigms

Fiammetta Namer,[a] Nabil Hathout[b]

[a] UMR 7118, ATILF, Université de Lorraine & CNRS, France
[b] UMR 5263, CLLE, Université de Toulouse Jean-Jaurès & CNRS, France

Abstract

In this article, we trace the genesis of the French derivational database Démonette and show how its architecture and content stem from recent theoretical developments in derivational morphology and from user needs. The development of this large-scale resource began a year ago as part of the Demonext project. Its conception is grounded in a theoretical approach where the lexemes are connected by derivational relations within derivational families which in turn fit into paradigms. More precisely, Démonette is a partial implementation of ParaDis, a paradigmatic model of morphological representation designed for the description of regular processes and of form-meaning discrepancies. The article focuses on the principles that govern the morphological, structural and semantic encoding of morphologically complex lexemes in Démonette and illustrates the range of form-meaning discrepancies with a variety of examples of non-canonical word formations.

1. Introduction

Démonette is a large-scale derivational database of French developed as part of the Demonext project funded by the French National Research Agency (ANR-17-CE23-0005). The project started in 2018. Its main goal is the description of 366,000 morphological relations covering a large range of processes that includes conversion, prefixation and suffixation. About 120 affixes are used in these derivations. They include suffixes like *-ard*, *-ariat*, *-at*, *-âtre*, *-el*, *-aie*, *-iser*, *-erie*, *-esque*, *-esse*, *-eur*, *-eux*, *-iste*, and prefixes like *a-*, *anti-*, *bi-*, *co-*, *contre-*, *dé-*, *é-*, *extra-*, *hyper-*, *hypo-*, *in-*, *infra-*, *inter-*. Démonette's descriptions primarily come from existing reliable resources created and

© 2020 PBML. Distributed under CC BY-NC-ND. Corresponding author: fiammetta.namer@univ-lorraine.fr
Cite as: Fiammetta Namer, Nabil Hathout. ParaDis and Démonette. From Theory to Resources for Derivational Paradigms. The Prague Bulletin of Mathematical Linguistics No. 114, 2020, pp. 5–33.
doi: 10.14712/00326585.001.

distributed as part of various academic works such as PhD projects. The data extracted from these databases is reanalysed in order to fit in the Démonette format and complemented when some features are missing. New entries extracted from machine readable dictionaries will also be added.

Démonette's entries are morphological relations between two lexemes (L_1, L_2) described by morphological, formal, categorical and semantic features. The database is highly redundant by design in order to be flexible enough and to have the capability to represent the many non-canonical morphological relations that occur in Word Formation (WF) of many languages. It is based on the theoretical principles that govern ParaDis, a paradigmatic model of derivational morphology where the multidimensional structure of the lexeme is generalized. ParaDis is based on two fundamental structures: derivational families (networks of lexemes) and paradigms (aligned sets of families). The article details the motivation of these theoretical assumptions and explains how they are implemented.

The remainder of the paper is organized as follows. In Section 2, we present an overview of some existing derivational databases (DDBs) that were created for different European languages. The theoretical background and motivation of the paradigmatic approach adopted in Démonette are then discussed in Section 3 with a focus on ParaDis and on the representation of various non-canonical derivational phenomena in this model. In Section 4, we first detail the structure of Démonette and the way the morphological, categorical and semantic features are encoded, and then show the capability of the proposed formalism to represent a number of non-canonical derivations that occur in French.

2. Resources and tools in morphology

Morphological analysis is part of the initial pre-processing task in many natural language processing (NLP) systems. The morphological analyzers they use are often based on machine learning and statistical methods. Words are decomposed into morphemes in order to compensate for the limitations of lexicons. Let us mention systems like Linguistica (Goldsmith, 2001), Morfessor (Creutz and Lagus, 2005), or, more recently, Cotterell and Schütze (2017)'s models. These systems may be used for any language provided that enough training data is available, however they are more effective for concatenative morphology languages such as English, German and French. Morphological analysis may also be carried out by symbolic (rule-based) systems developed by linguists; for a panorama, see (Bernhard et al., 2011).

Morphological parsers can be replaced or supplemented in the NLP pipeline by large enough lexical resources containing derivational annotations if their features are sufficiently rich and varied. However, very few such resources exist for most lan-

guages[1]. One of the firsts is CELEX (Baayen et al., 1995), which describes the phonetic, inflectional, morpho-syntactic, derivational and statistical properties of 216,775 lemmas of Dutch, English and German. The entries are extracted from dictionaries and corpora (news and literature). A more recent English DDB is CatVar (Habash and Dorr, 2003) which includes 100,000 lexemes grouped in subfamilies. A similar resource is DerivBase (Zeller et al., 2013) which describes 215,000 German lexical entries gathered in semantically motivated derivational families. Two other DDBs have recently been developed for German from CELEX, namely DErivCelex (Shafaei et al., 2017) and Morphological Treebank (Steiner and Ruppenhofer, 2018). Several DDBs have also been created for Slavic languages like CroDeriv (Šojat et al., 2014) for Croatian and DeriNet (Žabokrtský et al., 2016; Ševčíková et al., 2017; Vidra et al., 2019) for Czech. DeriNet is a lexical network that captures core word-formation relations connecting around 970,000 Czech lexemes. Derivational relations between verbs and some of their nominal derivatives are described in version 3.0 of Princeton (English) WordNet (Fellbaum et al., 2009) which provides a semantic characterization of the noun with respect to the verb. *Employer*$_{\mathrm{N}}$ is for instance described as the "agent" of *employ*$_{\mathrm{V}}$. DDBs have also been developed for Romance languages, mainly French, Italian and Latin. DerIvaTario is a derivational dictionary of Italian (Talamo et al., 2016) which provides descriptions based on strong hypotheses regarding allomorphy and suppletion. For instance, *bellicoso* 'bellicose' is analyzed as a derivative of *guerra* 'war'. Word Formation Latin (WFL) (Litta et al., 2016) is a derivational morphology resource for Classical Latin, where the lemmas (i.e., the non-inflected words) are decomposed into their formative components, and relations between the lemmas are identified by Word Formation Rules (WFRs). WFL contains 69,682 lemmas.

Few resources also exist for French. The JeuxDeMots platform (Lafourcade and Joubert, 2008), a serious game, has created a large coverage lexical network where the words are connected by semantic relations. These relations are inspired by the lexical function formalism (Mel'čuk, 1996). Some of them are derivational. JeuxDeMots being a crowd-sourced resource, the accuracy of the related pairs of words proposed by the players increases with the number of identical answers. In 10 years, the size of this resource has reached 270 million relations (pairs of words) that instantiate 150 different lexical functions. It connects 3.5 million words and expressions.

However, French still lacks true large-scale resources primarily aimed at the description of derivational morphology. To fill this gap, we developed a prototype database called Démonette$_{\mathrm{V1}}$ (Hathout and Namer, 2014b, 2016) from 2011 to 2017. Démonette$_{\mathrm{V1}}$ describes 73,233 derivational families made up of a verb, its agent and action noun derivatives and its modality adjective. Three objectives were pursued: (*i*)

[1]In his exhaustive review, Kyjánek (2018) proposes a typology of the structures and coverage of 30 derivational resources for Romance (including Latin), Germanic and Slavic languages and provides a complete list of the main existing DDBs and resources that contain derivational annotations. The reader may refer to this report which is far more complete than the present overview.

use the DériF morphological analyzer (Namer, 2009, 2013) to produce a resource made up of derivational relations between pairs of lexemes L_1 and L_2, labelled with linguistically grounded features, including semantic annotations; (*ii*) complement these $L_1 \rightarrow L_2$ derivations by indirect relations between members of the same derivational family extracted from the Morphonette lexicon (Hathout, 2009); (*iii*) define an extensible and redundant architecture which can be fed by heterogeneous morphological resources.

The design of the current Démonette database (Section 4) is based on the experience gained during the development of Démonette_{V1}. The aim of the second version is to produce a resource which provides descriptions (morphological, phonological, categorial, distributional, and especially semantic) that may be useful for NLP, but also serve as a reference for several audiences including research in morphology, teaching, language and speech therapy practice. The structure of the database must be flexible enough to allow for a (semi-)automatic acquisition of morphological descriptions from existing resources. It must also be able to include any non canonical formation or any additional derivation (affixation, conversion and even composition). To this purpose, the architecture of the database is based on theoretical principles that ensure a uniform representation of regular derivation (words where meaning and form mirror each other) and non-canonical derivation which infringe form-meaning compositionality. Démonette implements the principles of ParaDis, a model which borrows from lexeme-based and paradigm-based approaches to WF (Section 3).

3. Démonette's theoretical background

Démonette is based on two fundamental principles: (*i*) the adoption of the lexeme as unit of analysis, and (*ii*) the organisation of the morphological lexicon into paradigms. These principles have independently contributed to recent evolution of morphological theories, and influenced the content and organization of many derivational resources. This section briefly describes how these major modifications came about.

3.1. Morphemes and the form-meaning non-compositionality

The morpheme, conceived as the minimal bi-faced unit of meaning and form, is the descriptive and analytic unit morphologists have used in the so-called morpheme-based morphological traditions, whether concatenative (Item and Arrangement) or functional (Item and Process) (Hockett, 1954). Morpheme-based approaches have been adopted in many morphological analyzers and resources, including CELEX (Baayen et al., 1995), DerIvaTario (Talamo et al., 2016), CroDeriV (Šojat et al., 2014), the Morphological Treebank (Steiner and Ruppenhofer, 2018), and the first version of Word Formation Latin (Litta et al., 2016).

The main advantage of morpheme-based approaches is their simplicity and their capacity to describe all morphologically complex words by means of a small set of minimal units. Their integration into broader NLP systems is therefore very easy. In this approach, morphological rules handle only one type of unit, morphemes; they yield head-argument structures similar in nature to the outputs of syntactic rewriting rules: affixes are heads that select either lexical roots, or combinations of morphemes produced by other rules. The consequence of this similarity is that syntactic and morphological analysis and generation can be performed by a uniform grammatical system that operates on a reduced lexicon only made up of morphemes. However, this efficiency comes at a high cost. Morpheme-based morphology suffers from well-known limitations that have been widely discussed in the literature (among others, see (Aronoff, 1976; Anderson, 1992; Fradin, 2003); a recent, in-depth review is given in (Blevins, 2016)). The most significant drawback is the rigidity of the morpheme because it requires all form to be associated with a meaning and vice-versa. With such a strong constraint, the analysis for non canonical derivation processes (Corbett, 2010) becomes far too complex. Morphemes also prove unfit for the description of non-concatenative morphology languages such as templatic morphology in Semitic languages, tonal or stress shifting systems, etc. This takes away all interest in morpheme. Table 1 illustrates some of these limitations with English, French and Italian examples; similar formations exist in many other European languages including Spanish and German.

	WF	Lang.	Lexeme$_1$	Lexeme$_2$
a.	conversion	eng	*nurse*$_{\mathrm{N}}$	*nurse*$_{\mathrm{V}}$
b.	parasynthesis	fra	*banque*$_{\mathrm{N}}$ ‘bank’	*interbancaire*$_{\mathrm{A}}$ ‘between banks’
c.	parasynthesis	eng	*departement*$_{\mathrm{N}}$	*interdepartemental*$_{\mathrm{A}}$
d.	affix replacement	eng	*fascism*$_{\mathrm{N}}$	*fascist*$_{\mathrm{N}}$
e.	polysemy	fra	*porter*$_{\mathrm{V}}$ ‘carry’	*porteur*$_{\mathrm{Nm,[hum]\ or\ [artif]}}$ ‘carrier’
f.	synonymy	ita	*compatto*$_{\mathrm{A}}$ ‘compact’	*scompattare*$_{\mathrm{V}}$ or *decompattare*$_{\mathrm{V}}$ ‘uncompact’
g.	back-formation	eng	*vivisect*$_{\mathrm{V}}$ ‘perform vivisection’	*vivisection*$_{\mathrm{N}}$ ‘vivisection’

Table 1. Examples of meaning-form discrepancies in English, French and Italian derivational relations

The first example in Table 1(a) is a *zero affixation* or *conversion* (Tribout, 2012), also often called zero-morpheme (Dahl and Fábregas, 2018); it is characterized by Hathout and Namer (2014a) as a case of “formal under-marking” of the derivative with respect to its base since the form of the verb is identical to the form of the noun whereas the

semantic content of the verb is more complex: the verb *nurse* can be paraphrased as "act as a nurse", which includes the semantic content of the nominal homonym "nurse"; the predicative meaning "acting as" has no formal realization.

Another well-known type of meaning-form asymmetry illustrated in Table 1(b, c) is the so-called *parasynthetic* derivation; for recent overviews, see (Hathout and Namer, 2018; Iacobini, 2020). For instance, in Table 1(c), the *-al* suffix does not contribute to the meaning of *interdepartmental*, which only combines the meaning of the noun *department* and of the prefix *inter-*: "between departments". Likewise, in Table 1(b), the *-aire* suffix does not intervene in the construction the semantic content of *interbancaire* "between banks" which is derived by *inter-* prefixation, from the semantic content of the noun *banque*. At first sight, the prefix *inter-* is responsible of the semantic operation ("between Xs") and the suffix *-aire* of the change of categories (N→A). However, this analysis is challenged by the existence of unsuffixed adjectives like *interbank*, *interbirth*, *intercategory*, *interdeparment*, *interfamily* in English or *interbanques*, *interdépartements*, *intercellules*, *interatomes* in French. For instance, *interdepartment communication* can be found in English, or *transactions interbanques* "interbank transactions" in French. In other words, in Table 1(b) and 1(c), the prefix *inter-* assigns the derived words to their semantic class and grammatical category, whereas the suffix plays no role in the construction. Therefore, these derivatives are considered as "over-marked" in Hathout and Namer (2014a) because one of their formal elements does not have any semantic or categorial contribution.

Table 1(g) illustrates a similar case. On the semantic level, the verb could be considered as derived from the noun, since it has a more complex content than the noun. On the other hand, *vivisection* is 10 times more frequent than *vivisect* (10 times more Google hits) and its first occurrence is older: according to the Oxford dictionary, for example, the noun was in use at the beginning of the 18th century, whereas the verb's first occurrence dates back to mid 19th century. Moreover, *vivisect* means "perform a vivisection". In other words, *vivisect* is under-marked twice: the additional meaning in *vivisect* does not have a formal counterpart and there is no meaning associated with the *-ion* suffix in *vivisection*. This so-called "affix substraction" (Manova, 2011) is also known as *back-formation* (Becker, 1994).

The derivational relation between pairs of lexemes like *fascism*/*fascist* in Table 1(d) is analyzed as an *affix replacement* (Booij and Masini, 2015): Lexeme$_2$ is coined by replacing the *-ism* suffix in Lexeme$_1$ by *-ist* and vice-versa. Therefore, the two lexemes are "under- and over-marked" with respect to one another.

Other non-canonical derivations involve processes that produce two series of words that have the same form but different meanings or with different forms but the same meaning. In the first case (absence of formal markdown), the derivatives are *polysemous* as in Table 1(e) where French *-eur* suffixed nouns can denote humans and artifacts. In the second case, the derivatives are *synonymous*. This *morphological variation* results from a rivalry or *competition* between derivational processes which apply to the same base. In the Italian example in Table 1(f), the prefixes *s-* and *de-* compete to

form deadjectival verbs (Todaro, 2017). When applied to the same adjective *compatto*, they produce two synonymous verbs, *scompattare* and *decompattare*. This absence of semantic markdown can be regarded as a derivational equivalent of Thornton (2012)'s notion of *overabundance*.

3.2. Lexemes, and non-binary or non-oriented rules

The shift from morpheme to *lexeme* solves several problems that arise from meaning non-compositionality. Unlike the morpheme, the lexeme is not a concrete minimal unit. It is actually an abstract object (i.e. a noninflected word, in the simplest cases) that records the common properties of the inflectional paradigm it stands for, in the form of an autonomous three-dimensional structure: (*i*) a set of phonological form (or stems); (*ii*) a part-of-speech; (*iii*) a meaning. In this framework, *word* (or *lexeme*) *formation rules* (WFRs) are oriented relations between two schemata. Each of these schemata specifies the constraints the lexemes must meet in order to enter the relation and to activate the WFR. For instance, the English WFRs in the first column of Table 2 derive relation adjectives from nouns by suffixation in *-al* (Table 2(1a)) and *-ic* (Table 2(1b)); Table 3(1) presents the English WFR that converts nouns to similative verbs. The WFR states that the input nouns must denote human beings and that the output verbs are transitive. The derivational relations *government*/*governmental* in Table 2(2a), *atom*/*atomic* in Table 2(2b) and *nurse*/*nurse* in Table 3(2) respectively instantiate the WFR in Table 2(1a), Table 2(1b) and Table 3(1).

Because WFRs apply independently and simultaneously to all three levels of description (formal, categorial and semantic), more than one formal exponent can be associated with one semantic type of derivatives (like the competing affixes *-al* and *-ic* in Table 2). Similarly, a category-shifting process can be realized without any formal change as in Table 3. Conversely, one formal exponent can be associated with more than one semantic category of derivatives: for example, denominal adjectives of material, like *wooden* and deadjectival causative verbs like *blacken* are suffixed with the same *-en* exponent.

As we said, many problems illustrated in Table 1 are solved by the shift from morpheme to lexeme. A conversion like in Table 1(a) simply modifies the semantic content and the part-of-speech but leaves the formal content unchanged as shown in Table 3; similarly, a polysemous affixation as in Table 1(e) involves two distinct WFRs, one for humans and the other for artifacts. However, these WFRs are identical on the formal level: they use the same formal exponent to derive different semantic contents. Conversely, synonymy (Table 1(f)) corresponds to cases where two (or more than two) different WFRs apply to the same input lexeme (e.g. *compatto*) and produce two (or more than two) different formal realizations associated with the same derived semantic content.

However, some problems remain because WFRs are abstractions of oriented relations designed to connect derived words to their bases. They are for instance unfit for

	(1) WFR		(2) Example	
(a)	/X/ N '@'	→ /Xl/ A 'pertaining to @'	/ˈgʌvənmənt/ N 'government'	→ /ˌgʌvənˈməntl/ A 'pertaining to the government'
(b)	/X/ N '@'	→ /Xɪk/ A 'pertaining to @'	/ˈætəm/ N 'atom'	→ /əˈtɒmɪk/ A 'pertaining to the atom'

*Table 2. Two N→A*suf *Word Formation Rules in English*

(1) WFR		(2) Example	
/X/ N '@$_{+hum}$'	→ /X/ V$_{+transitive}$ 'act as a @'	/nɜːs/ N 'nurse'	→ /nɜːs/ V 'act as a nurse'

Table 3. N→V Word Formation Rule in English

the description of non-oriented and indirect derivational relations like affix replacement (Table 1(d)). Likewise, back-formation (Table 1(g)) cannot be represented by means of WFRs because the formal and semantic parts of the relation have opposite orientations, nor are they able to describe parasynthetic derivation like in (Table 1(b, c)). In this case, the limitation does not result from the orientation of the WFRs, but from the fact that these derivatives are produced by a ternary WF device. More specifically, classical WFRs cannot predict the value of the supernumerary suffix mark nor explain the diversity of these suffixes, as illustrated in Table 4. The adjectives *inter-banc**aire***, *intercellul**aire***, *interocéan**ique***, *interethn**ique***, *intertrib**al*** or *interparoissi**al*** all describe a spatial interval between two or more concrete entities ('between several X') where the noun X is, respectively, *banque*, *cellule*, *océan*, *ethnie*, *tribu* and *paroisse*. Prefixation in *inter-* may therefore involve at least three different suffix values (*-aire*, *-ique* and *-al*) but this value cannot be deduced from the form nor the meaning of the base. In other words, the adjectives in Table 4 cannot be properly analyzed without an access to the set of all the lexemes derivationally related to the base noun, as we will see below.

X_N	**between several** X_N	X_N	**between several** X_N
banque 'bank'	*interbancaire*	*cellule* 'cell'	*intercellulaire*
océan 'ocean'	*interocéanique*	*ethnie* 'ethny'	*interethnique*
tribu 'tribe'	*intertribal*	*paroisse* 'parish'	*interparoissial*

Table 4. Examples of the (X, interXsuf) noun-to-adjective relation in French

3.3. Paradigms, and partially motivated relations

Derivational paradigms solve most of the above-mentioned limitations raised by the lexeme-based approaches (for a panorama, see Štekauer (2014)) because paradigmatic relations are not necessarily binary nor are they oriented (base → derivative). In a paradigmatic framework (Bonami and Strnadová, 2019), the central unit is the *derivational family*, i.e. a structured set of lexemes[2] whose forms and meanings depend on each other. More specifically, all the members of a derivational family are interconnected just like all the inflected forms of a lexeme. Figure 1 adapted from (Bonami and Strnadová, 2019) presents a paradigm of four families made up of a verb and three derivatives (e.g. *advertise*, *advertiser*, *advertisement*, *advertisee*).

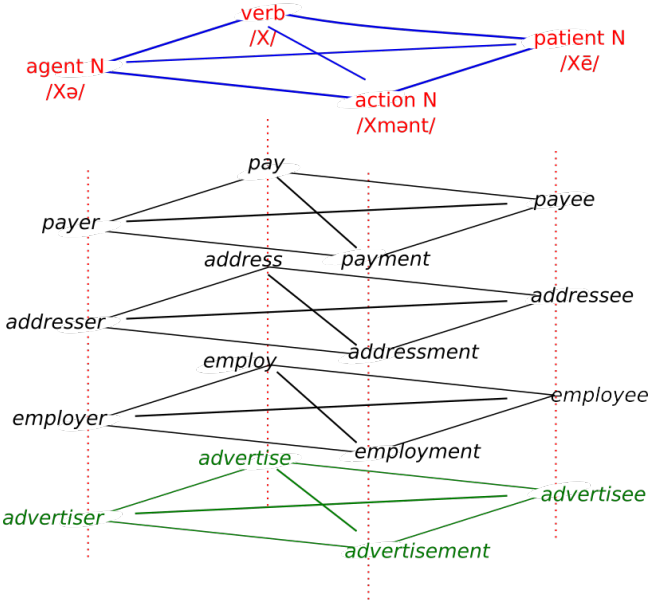

Figure 1. Regular paradigm

[2]However, the notions of paradigm and of lexeme are independent.

Two families F_1 and F_2 belong to the same paradigm when they line up so that members of the same rank or position in F_1 and F_2 are in the same form and meaning relations with the other members of their family. The aligned lexemes belong to the same *derivational series* (Hathout, 2011). For instance, the nouns *payee*, *addressee*, *employee*, *advertisee* in Figure 1 form a derivational series. A derivational paradigm can then be defined as a set of aligned families as illustrated in the lower part of Figure 1 which could be seen as a concrete paradigm in the sense of (McCarthy and Prince, 1993). The corresponding abstract paradigm is given in the upper part of Figure 1. An abstract paradigm is a network connecting the descriptions of the four derivational series.

A paradigm may contain some incomplete families, that is, families where some lexemes are missing with respect to other more complete families. Aligned incomplete families form sub-paradigms. For example, the paradigm in Figure 1 could be complemented by the 3-members family (*refuse*, *refuser*, *refusee*) which belongs to the sub-paradigm (verb-/X/, agent_N-/Xə/, patient_N-/Xiː/), but where the *-ment* action noun is missing.

Paradigm-based frameworks present two major advantages: flexibility and completeness. They are flexible because they do not only consider oriented base→derivative relations, and complete because their fundamental units are the derivational families. These two properties enable paradigm-based models to take into account affix replacement (Table 1(d)) and back-formation (Table1(g)) in a straightforward way.

In a paradigmatic approach of WF, the relations between the members of a derivational family are all represented in the same way, as non-oriented schemata, be the relations direct (base→derivative) or not. For example, the schema (1) describes the relation between *fascist* and *fascism* of (Table 1(d)): the "@1" and "@2" variables stand for the semantic content of *fascist* and *fascism* respectively, and X stands for the sequence /fæʃ/ they have in common. The mutual motivation of the two nouns can be expressed by a cross-definition of their semantic content: *fascism* is defined as the "ideology defended by a fascist" and *fascism* as a "follower of fascism"[3]. The *fascism*↔*fascist* pair is a partial family that fits in a larger paradigm represented in Table 5. The triplets (Table 5(a, b)) connect a noun or a proper name referring to an entity X, a noun of ideology (Xism) that values that entity, and a human noun (Xist) denoting a person supporting that ideology. Bochner (1993) represents these paradigmatic relations in the theoretical framework of the *Cumulative Patterns* as a ternary schema as in (2)[4].

[3]Booij and Masini (2015) propose a slighly different way to formalize cross-formation patterns by means of so-called "second order schemata".

[4]Other theoretical frameworks have been proposed to represent paradigms in derivation by Koenig (1999); Booij (2010); Spencer (2013); Antoniova and Štekauer (2015) to only cite a few.

	X: valued entity	Xism: ideology	Xist: follower
a.	*Calvin*	*calvinism*	*calvinist*
b.	*race*	*racism*	*racist*
c.	–	*fascism*	*fascist*

Table 5. (X, Xist, Xism) paradigm in English

$$(1)\quad \left[\begin{array}{l} /\mathrm{X}\textsc{ɪst}/ \\ \mathrm{N} \\ \text{@1:'follower of @2'} \end{array}\right] \leftrightarrow \left[\begin{array}{l} /\mathrm{X}\textsc{ɪzm}/ \\ \mathrm{N} \\ \text{@2: 'ideology defended by @1'} \end{array}\right]$$

$$(2)\quad \left\{ \left[\begin{array}{l} /\mathrm{X}\textsc{ɪst}/ \\ \mathrm{N} \\ \text{@1:'follower of @2,} \\ \text{endorsing @3'} \end{array}\right], \left[\begin{array}{l} /\mathrm{X}\textsc{ɪzm}/ \\ \mathrm{N} \\ \text{@2: 'ideology} \\ \text{defended by @1,} \\ \text{promoting @3'} \end{array}\right], \left[\begin{array}{l} /\mathrm{X}/ \\ \text{PrN or N} \\ \text{@3: 'entity} \\ \text{promoted by @2,} \\ \text{endorsed by @1'} \end{array}\right] \right\}$$

The description of back-formation (Table 1(g)) is also straightforward, because the paradigmatic relation (3) is not oriented. On the formal level, it connects *vivisect* to *vivisection* where X stands for /ˌvɪvɪˈsek/. On the semantic level, it connects *vivisection* to *vivisect* through the definition of '@1'(the meaning of the verb) as being derived from '@2' (the meaning of the dynamic noun).

$$(3)\quad \left[\begin{array}{l} /\mathrm{Xt}/ \\ \mathrm{V} \\ \text{@1:'perform @2'} \end{array}\right] \leftrightarrow \left[\begin{array}{l} /\mathrm{X}\textsc{ʃən}/ \\ \mathrm{N}_{+\mathrm{dyn}} \\ \text{@2} \end{array}\right]$$

Nevertheless, the questions raised by the parasynthetic derivations in Table 1(b,c) still remain to be answered. They are illustrated in Table 6 which repeats and extends Table 4 where we have seen that all the prefixed adjectives contain a suffix, that its value is variable and that it cannot be predicted from the form nor meaning of the noun; we have remarked that the suffix does not contribute to the meaning of the prefixed adjective, which is derived directly from the meaning of the noun. For all these reasons, we have said that the prefixed adjective is over-marked with respect to its base noun.

However, the over-marking is not arbitrary, as illustrated in Table 6: the suffix that appears in the prefixed adjective (column 3) is always the same as the suffix of the relational adjective (column 2) from the noun (column 1). In other words, the form of the prefixed adjective is derived from the relational adjective of the noun while its meaning is derived from the meaning of the noun. Their construction uses both the semantic properties of X_N and the formal properties of $Xsuf_A$. Therefore, the *inter*$Xsuf_A$ adjectives have two bases, one semantic (the noun X_N) and one formal (the relational adjective $Xsuf_A$).

X_N	$Xsuf_A$: 'of X'	$interXsuf_A$: 'between several Xs'
banque	*bancaire*	*interbancaire*
cellule	*cellulaire*	*intercellulaire*
tribu	*tribal*	*intertribal*
paroisse	*paroissial*	*interparoissial*
océan	*océanique*	*interocéanique*
ethnie	*ethnique*	*interethnique*
corail 'coral'	*corallien*	*intercorallien*
bactérie 'bactery'	*bactérien*	*interbactérien*

Table 6. (X, Xsuf*,* interXsuf*) paradigm in French.*

The classical paradigmatic approaches mentioned above consider that WF takes place in the derivational families. However, they are not able to handle the parasynthetic derivatives because they are designed to describe derivational relations where the three dimensions of the lexeme (form, category and meaning) co-vary. To overcome this limitation, we need a model where the semantic and formal relations are described separately, as they are in ParaDis.

3.4. ParaDis

Asymmetric formations like the ones in Table 1(b, c) are far from exceptional. They occur in French and in many European languages and concern a large portion of the denominal prefixed adjectives. In French, these adjectives describe spatial relations (*inter-*, *intra-*, *sous-*, *sur-*, etc.), temporal relations (*pré-*, *post-*, *anté-*, etc.), opposition (*anti-*), support (*pro-*), quantification (*mono-*, *bi-*, *pluri-*, etc.) and many others. They also concern denominal verbs like *lieu* 'place' → *localiser* 'localize'; for a full overview, see (Hathout and Namer, 2014a).

In order to account for these formations, we need a model that transposes the main contribution of lexeme-based morphology (the independent formal, categorial and semantic levels of representations) to the paradigmatic organization of the lexicon. The model must combine a morpho-phonological structure where the form of an $interXsuf_A$ adjective is connected to the form of the corresponding $Xsuf_A$ with a morpho-semantic structure where the meaning of $interXsuf_A$ is related to the meaning of X_N.

This description can be framed in the theoretical framework ParaDis "Paradigms vs Discrepancies" (Hathout and Namer, 2018) which generalize the three levels structure of the lexicon to the derivational paradigms. Our assumption is that derivational morphology is paradigmatic because the morpho-semantic regularities, the morpho-categorial regularities and the morpho-formal regularities are paradigmatic.

In other words, the organization of the derivational paradigms is extended to the semantic, categorial and formal levels of representation. In ParaDis, a morphological paradigm is an abstract combination of a morpho-formal paradigm, a morpho-categorial paradigm and a morpho-semantic paradigm just as a lexeme is the abstract combination of a formal, a categorial and a semantic descriptions. The combination of the three paradigms is obtained by mapping each of them to the morphological paradigm. In this framework, the formal, categorial and semantic levels are independent in the sense that there are not directly connected. This independence is key to the description of the asymmetric formations like *interbancaire*, as shown in Figure 2. For the sake of readability, the categorial and the semantic levels have been merged in the remainder of this article.

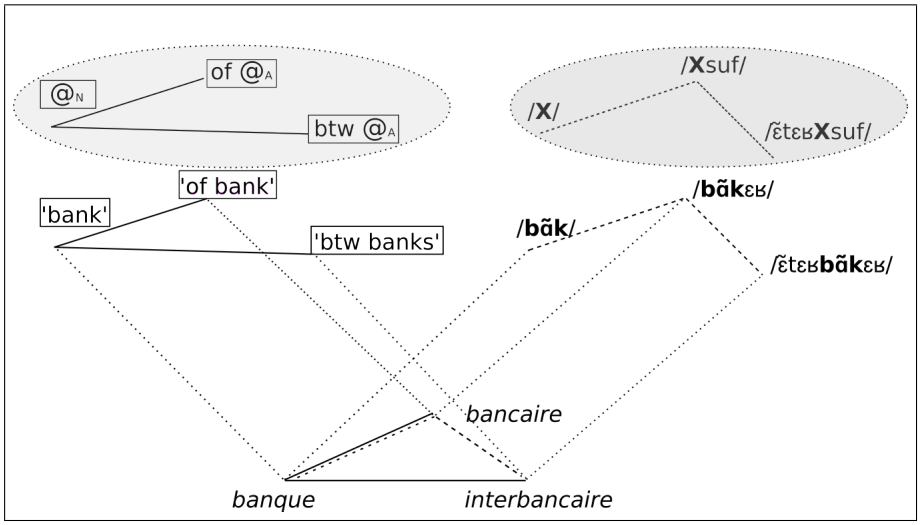

Figure 2. ParaDis. Representation of the (X_N, $Xsuf_A$, $interXsuf_A$) asymmetrical paradigm

In Figure 2, the gray oval on the right represents a formal abstract paradigm defined as a network of formal series. This abstract paradigm is an abstraction of a concrete formal paradigm defined as an alignment of formal families. It is represented in the figure by a single formal family (i.e. the network that is just below), the one that contains the form of the prefixed adjective *interbancaire*. The other formal families of the concrete formal paradigm are omitted. In the formal family, the vertices are phonological representations and the edges (dashed lines) describe the formal motivation that hold between these representations.

Similarly, the gray oval on the left represents a semantic abstract paradigm and the network just below the semantic family of the meaning of the prefixed adjective. The nodes in this graph represent meanings (morpho-semantic values) and the edges (solid lines) describe how they are related to the other meanings contained in the family.

The graphs of the formal and the semantic families are incomplete and different. In the semantic graph, the meanings 'btw banks' and 'of banks' cannot be deductible one from the other; these nodes are not connected. Likewise, /bãk/ and /ɛ̃tɛʁbãkɛʁ/ are not connected in the formal family because the ending (i.e. suffix) of the latter cannot be predicted from the former.

The formal and semantic concrete paradigms are in correspondence (dotted lines) with a morphological paradigm represented in the lower part of the figure by one of the morphological families it contains, namely the family of *interbancaire*. As mentioned above, morphological paradigms are alignments of morphological families and morphological families are connected graphs of lexemes.

When applied to the (*banque*, *bancaire*, *interbancaire*) family, the projection of the formal and semantic families on the morphological family results in three types of relations. The relation between *banque* and *bancaire* is regular (two lines, one solid and one dashed): it inherits the semantic motivation from the semantic family, and the formal motivation from the formal family. On the other hand, the relation between *bancaire* and *interbancaire* has only a formal motivation (dashed line) and the relation between *banque* and *interbancaire* has only a semantic motivation (solid line). The other families of Table 6 are analyzed in the same way.

X_N '@'	$Xsuf_A$ 'of @'	$pluriXsuf_A$ 'with more than one @'	$multiXsuf_A$
atome 'atom'	*atomique*	*pluriatomique*	*multiatomique*
cellule 'cell'	*cellulaire*	*pluricellulaire*	*multicellulaire*
clone 'clone'	*clonal*	*pluriclonal*	*multiclonal*
os 'bone'	*osseux*	*pluriosseux*	*multiosseux*

Table 7. (X_N, Xsuf$_A$, pluriXsuf$_A$, multiXsuf$_A$) families in French

The above analysis can be extended to the French families of the form (X_N, $Xsuf_A$, $pluriXsuf_A$, $multiXsuf_A$) illustrated in Table 7. In these families both prefixes express plurality; *multicellulaire* and *pluricellulaire* are synonymous; they mean 'with more than one cell'. These families raise two meaning-form issues. First, the prefixed adjectives (in columns 3 and 4) are always over-marked, and contain a semantically neutral suffix borrowed from the same relational adjective in column 2, whereas their meaning is directly computed from the semantic content '@' of the noun X in column 1. The

different values of the suffix (*-ique*, *-aire*, *-al*, *-eux*) reflect the competition that exists between these WF processes. The second issue is that the two prefixed adjectives are synonymous: with concrete nominal bases, the *pluri-* and *multi-* prefixes can be freely substituted one for the other (Amiot, 2005). Similar synonymous parasynthetic adjectives also exist in other Romance languages and in English. The combination of these mismatches results in the apparent irregularity of the derivational relations within the families in Table 7. The way they are analyzed in ParaDis is illustrated by Figure 3.

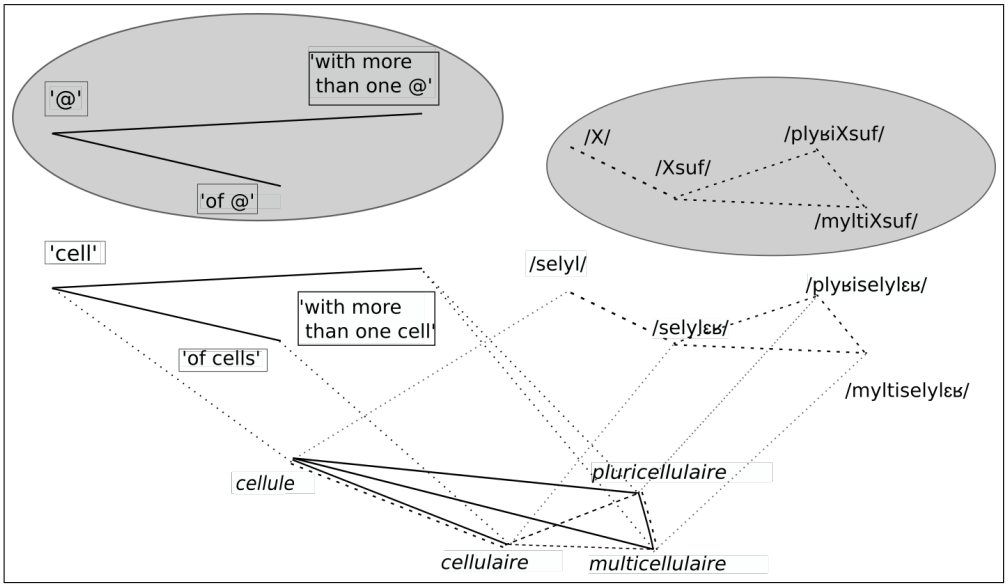

Figure 3. ParaDis. Representation of the X_N, $Xsuf_A$, $pluriXsuf_A$, $multiXsuf_A$*) asymmetrical paradigm*

Semantically, the families in Table 7 form a three-cells paradigm similar to the semantic paradigm in Figure 2: the entity '@' is connected to the relation 'of @' and the modifier 'with more than one @'. In this graph, the meanings of the relation and the modifier are not directly related. On the formal side (at the top right), /selylɛʁ/, /plyʁiselylɛʁ/ and /myltiselylɛʁ/ are connected because they are inter-predictible; /selylɛʁ/ depends on /selyl/; on the other hand, /plyʁiselylɛʁ/ and /myltiselylɛʁ/ cannot be predicted from /selyl/. The meaning-form asymmetry in the morphological families in Table 7 results from the differences between the formal and the semantic graphs. The two graphs have different sizes with four vertices for the formal network

but only three for the semantic one. Because the semantic graph is smaller, the difference in size expresses a regular synonymy: the meaning 'with more than one cell' corresponds to two distinct members in the morphological family (*multicellulaire* and *pluricellulaire*).

The next section shows how the main features of ParaDis are implemented in Démonette.

4. The **Démonette** derivational database

Démonette is a derivational database fully compatible with the principles presented in Section 3. It is able to uniformly represent the classical binary and oriented derivation processes and all the meaning-form mismatches illustrated in Table 1.

Démonette has an original structure: its entries are *derivational relations* between pairs of lexemes that belong to the same family. They are not limited to relations between a base and one of its derivatives and include back-formations, cross-formations, parasynthetical derivatives, etc. In addition to the initial 96,000 entries of $\text{Démonette}_{\text{V1}}$ (Hathout and Namer, 2014b; Namer et al., 2017), Démonette is fed by several existing derivational resources developed and validated by morphologists. These reliable resources contain detailed semantic and phonological descriptions. 183,000 entries will be added in this way. Most of them are direct relations corresponding to ca. 120 derivational processes: conversion; suffixation in *-ard*, *-ariat*, *-at*, *-âtre*, *-el*, *-aie*, *-iser*, *-erie*, *-esque*, *-esse*, *-eur*, *-eux*, *-iste*, etc.; prefixation in *a-*, *anti-*, *bi-*, *co-*, *contre-*, *dé-*, *é-*, *extra-*, *hyper-*, *hypo-*, *in-*, *infra-*, *inter-*, etc. The original base→derivative relations are cast into the Démonette's format and new information, new pairs and new lexemes are (semi-)automatically added when necessary.

In what follows, we present Démonette's general structure (§ 4.1). We then detail how the regular and irregular derivational relations are represented (§ 4.2), including polysemy, conversion, back-formation and cross-formation (see Table 1). We also show how synonymous and parasynthetic derivatives are represented in Démonette (§ 4.3).

4.1. Overview

Démonette implements the main features of ParaDis. Its structure is based on the following principles, some of which having been already implemented in $\text{Démonette}_{\text{V1}}$:

- a *record* or *entry* describes a relation between two lexemes that belong to a derivational family; is identified by a pair of lexemes;
- a *lexeme* that takes part in several relations will be described in as many records;
- an entry (L_1, L_2) contains the *description* of L_1, of L_2 and of relation that holds between them (that is, the derivational pattern that generalizes the relation (L_1, L_2));

L	PoS	Inflectional paradigm (Latinate root)	Ontological type	...
planter	V	plãt, plãtɔ̃, plãte, plãtɛ, ... (plãtat)	Dynamic Situation	...
planteuse	Nfem	plãtøz	Person|Artifact	...

Table 8. Démonette. Excerpt of the table of lexemes

- some *features* of a lexeme are independent of the relations it appears in. They include the standardized written form of the lexeme, its part-of-speech, its inflectional paradigm (in IPA format), a possible set of learned roots (e.g. the latinate root plãtat for the verb *planter* 'plant'; see Table 8), and its ontological type selected among the 25 WordNet Unique Beginners (UB) (Miller et al., 1990). The description of the lexemes are grouped into a **table of lexemes**. Table 8 presents an excerpt of the records of the verb $planter_{\text{V}}$ 'plant' and $planteuse_{\text{Nfem}}$ 'female planter' or 'instrument used to plant (trees)'. Derivational polysemy is described in the table of lexemes where the ambiguity between several related meanings are indicated by the symbol "|", meaning "or", as illustrated by the ontological type of *planteuse*;
- *relations* between lexemes are stored in a separate table, the **table of relations**;
- a *relation* (L_1, L_2) is defined by three independent sets of features: morphological (characterization of the morphological process connecting L_1 to L_2), formal (description of the formal variation between L_1 and L_2) and semantic (semantic category of the relation and glosses define L_1 and L_2 with respect to one another).

The remainder of the paper details the architecture of Démonette and focuses on the formal, structural and semantic features of the table of relations; readers can refer to (Namer et al., 2017) for a presentation of the morpho-phonological features. We also show how meaning-form discrepancies are taken into account and how this resource can provide a large-scale description of the paradigmatic organization of the morphologically complex lexicon.

4.2. (Almost) regular paradigms in Démonette

Démonette is a suitable tool for the representation of regular relations in word-formation: canonical derivation, cross- and back-formation, conversion, cf. Table 1. First, consider the relations between the lexemes of the families in Table 9. Theses families contain a verb (the predicate *laver* 'wash'), the corresponding iterative verb (the predicate *relaver* 're-wash'), the action noun derived from the two predicates (*lavage* 'washing' and *relavage* 'rewashing'), and an adjective expressing potentiality (*lavable* 'washable'). In French, action nouns may be derived by conversion (*découper* → *découpe*) or suffixation in *-age*, *-ment*, *-ion*, *-ure*, etc.; all these processes are in competi-

	X_V	$X(suf)_N$	reX_V	$reX(suf)_N$	$Xable_A$
a.	*laver* 'wash'	*lav**age***	*relaver*	*relav**age***	*lavable*
b.	*classer* 'rank'	*classe**ment***	*reclasser*	*reclasse**ment***	*classable*
c.	*planter* 'plant'	*plantat**ion***	*replanter*	*replantat**ion***	*plantable*
d.	*souder* 'weld'	*soud**ure***	*resouder*	*resoud**ure***	*soudable*
e.	*découper* 'cut (out)'	*découpe*	*redécouper*	*redécoupe*	*découpable*

Table 9. (X_V, $X(suf)_N$, reX_V, $reX(suf)_N$, $Xable_A$) families in French

tion. However, in each families in Table 9, the action nouns of the two predicates are always derived by the same formal process. All the derivational relations between the members of the families in Table 9 are regular because they all are formally and semantically motivated. They form complete oriented graphs. These graphs contain 20 edges and each of them is an entry in Démonette. We will see in § 4.2.1 and § 4.2.2 how the formal and the semantic features interact in order to represent different categories of regular WF relations illustrated by the family of *laver* (Table 9(a)).

4.2.1. Morphological features

Table 10 lists the morphological relations that hold between the members of the family of *laver* with their structural and morphological features. The relations in the other families in Table 9 are described in the same way; the formal aspect related to the conversion in Table 9(e) are discussed in Table 11. The morphological description of a relation (L_1, L_2) involves five features: `Orientation` and `Complexity` encode the structural properties of the relation; the values of $\texttt{Schema}_{L1}$ and $\texttt{Schema}_{L2}$ correspond to the morphological patterns of L_1 and $L_2$with respect to this relation; `Morph(ological) Match(ing)` combines $\texttt{Schema}_{L1}$ and $\texttt{Schema}_{L2}$.

- The `Orientation` of entry (L_1, L_2) indicates whether L_1 is an ancestor of L_2 (`a2d`; ancestor to descendant), whether L_2 is an ancestor of L_1 (`d2a`; descendant to ancestor) or whether the relation is `indirect`.
- `Complexity` describes the number of morphological steps needed to reach L_2 from L_1. When one lexeme is the base of the other, the value is `simple` (e.g. *laver* is the base of *lavage*). The value is also `simple` when L_1 and L_2 have a common base (e.g. *lavage* and *relaver* are both derived from *laver*). Notice that a derived word may have more than one base. For example, *relavage* is derived from *relaver* by suffixation in *-age* and from *lavage* by prefixation in *re-*. In both entries (*relavage*, *relaver*) and (*relavage*, *lavage*), `Orientation` is `a2d` and `Complexity` is `simple`. The value `complex` is used in all the other cases. A `complex` relation has a `a2d` or `d2a` orientation when it connects an ancestor and a descendant and involves at least two steps (e.g. (*relavage*, *laver*) is a two-steps relation where *laver*

L_1	L_2	$Schema_{L1}$	$Schema_{L2}$	Morph Match	Orientation	Complexity
laver	*lavage*	X	Xage	X/Xage	a2d	simple
laver	*relavage*	X	reXage	X/reXage	a2d	complex
laver	*relaver*	X	reX	X/reX	a2d	simple
laver	*lavable*	X	Xable	X/Xable	a2d	simple
lavage	*relavage*	X	reX	X/reX	a2d	simple
lavage	*relaver*	Xage	reX	Xage/reX	indirect	simple
lavage	*lavable*	Xage	Xable	Xage/Xable	indirect	simple
relaver	*relavage*	X	Xage	X/Xage	d2a	simple
relavage	*lavable*	reXage	Xable	reXage/Xable	indirect	complex
relaver	*lavable*	reX	Xable	reX/Xable	indirect	simple

Table 10. Démonette. Structural and formal features of the morphological relations that hold in the family of laver

is an ancestor of *relavage*). The relations are `complex, indirect` if they hold between two distant members and neither of them is a descendant or an ancestor of the other, e.g. (*relavage*, *lavable*).

- $Schema_{L1}$ and $Schema_{L2}$ describe the exponents of L_1 to L_2 in the relation that connects L_1 to L_2: X represents the sequence they have in common in this context. Therefore, the schemata are relation-dependent and vary with respect to the relation. For instance, *relavage* is annotated `Xage` with respect to *relaver* and `reXage` with respect to *lavable*.
- `Morph(ological) Match(ing)` is a concatenation of the values of $Schema_{L1}$ and $Schema_{L2}$. Two relations with identical `Morph(ological) Match(ing)` belong to the same morphological series regardless of the part of speech involved: the ($laver_V$, $lavage_{Nmas}$) and ($relaver_V$, $relavage_{Nmas}$) pairs share the value `X/Xage` and therefore belong to the same series; likewise, ($laver_V$, $relaver_V$) and ($lavage_{Nmas}$, $relavage_{Nmas}$) belong to the same series, identified by the `X/reX` value.

For the sake of space, the relations in Tables 10, 11 and 12 are listed in only one direction. Any entry (L_1, L_2) in Démonette has a symmetrical entry (L_2, L_1). In this entry, the values of $Schema_{L1}$, $Schema_{L2}$ are inverted; the value of `Morpho(logical) Match(ing)` is the mirror of that of (L_1, L_2); the value `a2d` becomes `d2a` and vice-versa for the feature `Orientation`; the other features are unchanged.

Verb-noun conversion (Table 11) can be described with the same five features. Since conversion does not involve exponents, the values of $Schema_{L1}$ and $Schema_{L2}$ are always `X` and `X` and the value of `Morphological Matching` is always `X/X`. The lack of exponent makes it impossible to decide which of the noun and the verb is the base; for a complete analysis of verb-noun conversion in French, see (Tribout, 2012). There-

fore the `Orientation` is non-documented (`nd`) for the conversions, as in Table 11(a,b). However, the `Orientation` may be known in two cases:

1. When the noun contains an exponent which shows that it results from a derivational process that cannot yield a verb, then the noun is the base of the conversion and the verb derives from it. For instance, in Table 11(c), the noun is a neoclassical compound, and in French, neoclassical compounding never produces verbs. Therefore, *hydrogéner*$_{\text{V}}$ derives from *hydrogène*$_{\text{Nmas}}$.
2. Symmetrically, when the verb contains a formal mark showing that it results from an affixation, and that this affixation cannot yield a noun, then the verb is the base. In Table 11(d), the intensive prefixation in *dé-* only produces verbs. Therefore, the noun *découpe* derives form the verb *découper*.

We can use the exact same features to describe cross-formation (Table 1(d)). The value `indirect` of `Orientation` indicates that both lexemes have exponents and that they are substituted one for the other. Table 12(a, b, c) describes the relations between the members of the (*race*, *racisme*, *raciste*) family, and Table 12(d) the (*fascisme*, *fasciste*) relation. In Table 12(c, d), the values of `Orientation`, `Complexity`, and `Morpho(logical) Match(ing)` are the same for the cross-formation relations (*fascisme, fasciste*) and (*racisme, raciste*). This shows that the incomplete family of (*fascisme, fasciste*) belongs to a sub-paradigm of the paradigm of (*race, raciste, racisme*).

Back-formation can also very easily be described by means of Démonette's features as in Table 13(a, b, c). The verbs (L_1) and the nouns (L_2) start with the same neoclassical components (*thermo-* 'heat', *hydro-* 'water' and *aéro-* 'air'). In addition, the nouns are suffixed by *-age*. Formally, the nouns are more complex than the corresponding verb as indicated by the value `X/Xage` for `Morph Match` which also describes regular suffixation in *-age* as in Table 13(d). However, for the back-formations of Table 13(a, b, c), the value of `Orientation` is `d2a` and not `a2d` as in regular derivations (Table 13(d)). This value expresses the fact that the verb is derived from the noun, for the same reason as with *hydrogène*$_{\text{Nmas}}$→*hydrogéner*$_{\text{V}}$ in Table 11(c): the nouns in Table 13(a, b, c) are formed by neoclassical compounding, like *collage*→*thermocollage* (Table 13(e)) and neoclassical compounding cannot not yield verbs in French. This means that the verbs in Table 13(a, b, c) are derived from the (formally more complex) nouns.

	L_1	L_2	Orientation	Complexity
a	*scier*$_{\text{V}}$ 'to saw'	*scie*$_{\text{Nfem}}$ 'saw'	nd	simple
b	*danser*$_{\text{V}}$ 'to dance'	*danse*$_{\text{Nfem}}$ 'dance'	nd	simple
c	*hydrogéner*$_{\text{V}}$ 'to hydrogenate'	*hydrogène*$_{\text{Nmas}}$ 'hydrogene'	d2a	simple
d	*découper*$_{\text{V}}$ 'to cut out'	*découpe*$_{\text{Nfem}}$ 'cut'	a2d	simple

Table 11. Démonette. Verb-noun conversion

	L_1	L_2	$Schema_{L1}$	$Schema_{L2}$	Morph Match	Orientation	Complexity
a.	*race*	*racisme*	X	Xisme	X/Xisme	a2d	simple
b.	*race*	*raciste*	X	Xiste	X/Xiste	a2d	simple
c.	*racisme*	*raciste*	Xisme	Xiste	Xisme/Xiste	indirect	simple
d.	*fascisme*	*fasciste*	Xisme	Xiste	Xisme/Xiste	indirect	simple

Table 12. Démonette. Cross formation: Structural and formal properties of race/raciste/racisme *and* fascisme/fasciste *families*

4.2.2. Semantic features

Démonette provides a semantic description for the relations where `Complexity =simple`. It includes the semantic type of the relation (`SemRel`), a gloss in natural language which defines L_1 and L_2 with respect to each other (`Concrete Definition`) and a generalization of this cross-definition (`Abstract Definition`). Table 14 presents examples of these semantic description.

The value of `SemRel` depends on a combination of features that describe L_1, L_2 and their relation: the ontological classes of L_1 and L_2; the parts-of-speech of L_1 and L_2; the `Orientation`, `Complexity` and `Morph(ological) Match(ing)` of the relation. Different combinations may correspond to the same value of `SemRel`.

The values for `SemRel` in Table 14 are `syn(onymy)`, `iter(ativity)` and `pot(entiality)`. The value is `syn(onymy)` for relations between a dynamic predicate and its derived action noun (when both denote dynamic situations (`Onto.Type=Dyn-Situation`) and when the morphological properties of the relation describe a direct base-derivative derivation rule, e.g. `Morph Match=X/Xage` as in Table 14(a, b)). `SemRel` equals `iter(ativity)` when the value of `Morph Match` involves an iterative prefixation like *re-* (`reX/X` in Table 14(c, d) or `reX/Xage` in Table 14(e)). Its value is `pot(entiality)` when L_1 or L_2 denotes a dynamic predicate and the other lexeme denotes an *-able* suffixed modifier, as in Table 14(f, g, h).

	L_1	L_2	Morph Match	Orientation	Complexity
a.	*thermocoller*	*thermocollage*	X/Xage	d2a	simple
b.	*hydromasser*	*hydromassage*	X/Xage	d2a	simple
c.	*aérosonder*	*aérosondage*	X/Xage	d2a	simple
d.	*coller*	*collage*	X/Xage	a2d	simple
e.	*thermocollage*	*collage*	thermoX/X	d2a	simple

Table 13. Démonette. Representation of Back Formation

	L_1 & Ont.Type$_{L1}$	L_2 & Ont.Type$_{L2}$	Morph Match	Sem Rel	Concrete Definition & Abstract Definition
a.	*laver*$_{V1}$ Dyn-Sit$_{V1}$	*lavage*$_{N2}$ Dyn-Sit$_{N2}$	X/Xage	syn	'laver$_{V1}$ sth is to perform lavage$_{N2}$ on it' 'Dyn-Sit$_{V1}$ sth is to perform Dyn-Sit$_{N2}$'
b.	*relaver*$_{V1}$ Dyn-Sit$_{V1}$	*relavage*$_{N2}$ Dyn-Sit$_{N2}$	X/Xage	syn	'relaver$_{V1}$ sth is to perform relavage$_{N2}$ on it' 'Dyn-Sit$_{V1}$ sth is to perform Dyn-Sit$_{N2}$'
c.	*laver*$_{V1}$ Dyn-Sit$_{V1}$	*relaver*$_{V2}$ Dyn-Sit$_{V2}$	X/reX	iter	'laver$_{V1}$ sth several times is to relaver$_{V2}$ it' 'Dyn-Sit$_{V1}$ sth several times is to Dyn-Sit$_{V2}$ it'
d.	*lavage*$_{N1}$ Dyn-Sit$_{N1}$	*relavage*$_{N2}$ Dyn-Sit$_{N2}$	X/reX	iter	'Perform several lavage$_{N1}$ is to perform relavage$_{N2}$' 'Perform several Dyn-Sit$_{N1}$ is to perform Dyn-Sit$_{N2}$'
e.	*lavage*$_{N1}$ Dyn-Sit$_{N1}$	*relaver*$_{V2}$ Dyn-Sit$_{V2}$	Xage/reX	iter	'Perform several lavage$_{N1}$ is to relaver$_{V2}$' 'Perform several Dyn-Sit$_{N1}$ is to Dyn-Sit$_{V2}$'
f.	*laver*$_{V1}$ Dyn-Sit$_{V1}$	*lavable*$_{A2}$ Mod$_{A2}$	X/Xable	pot	'One can laver$_{V1}$ sth if it is lavable$_{A2}$' 'One can Dyn-Sit$_{V1}$ sth if it is Mod$_{A2}$'
g.	*lavage*$_{N1}$ Act$_{N1}$	*lavable*$_{A2}$ Mod$_{A2}$	Xage/Xable	pot	'One can perform lavage$_{N1}$ on sth if it is lavable$_{A2}$' 'One can perform Dyn-Sit$_{N1}$ on sth if it is Mod$_{A2}$'
h.	*relaver*$_{V1}$ Dyn-Sit$_{V1}$	*lavable*$_{A2}$ Mod$_{A2}$	reX/Xable	pot	'One can relaver$_{V1}$ sth if it is lavable$_{A2}$ several times' 'One can Dyn-Sit$_{V1}$ sth if it is Mod$_{A2}$ several times'

Table 14. Démonette. Semantic features of the relations in the family of laver

The values of `Concrete Definition` are inspired by Frame Semantics tradition and especially *FrameNet*, its most popular implementation (Fillmore, 2006). The fundamental assumption is that people understand language through situations evoked in their mind by certain words. These representations are called frames, and involve the participants to the situation. Unlike frames, the situations described in the `Concrete Definition` glosses are derivationally relevant but may not be relevant cognitively; see (Sanacore et al., 2019).

The `Abstract Definitions` are generalizations of the `Concrete Definitions` where L_1 and L_2 are replaced by their ontological types. For instance, in Table 14(g), the `Concrete Definition` of (*lavage*$_{N1}$, *lavable*$_{A2}$) is 'One can perform *lavage*$_{N1}$ on something if it is *lavable*$_{A2}$'; in the corresponding `Abstract Definition`, *lavage*$_{N1}$ replaced by Dyn-Sit$_{N1}$ and *lavable*$_{A2}$ by Mod$_{A2}$. Derivational relations with the same `Abstract Definition` belong to the same semantic series like (*laver*, *lavage*) in Table 14(a) and (*relavage*, *relaver*) in Table 14(b)[5].

Table 15 presents an example of the description of rival WF processes in Démonette. The derivational relations listed in this Table are the same as in Table 9 (columns

[5]The semantic features of symmetrical pairs (L_1, L_2) and (L_2, L_1) are identical when their indexes are switched.

1 and 2). They involve competing WF processes because their descriptions are identical except for $\texttt{Schema}_{L1}$ or $\texttt{Schema}_{L2}$ (and consequently for `Morphological Matching`) and of course for `Concrete Definition` because the lexemes are different. The identical features are omitted: `Orientation=a2d`, `Complexity=simple`, `Onto.Type=Dyn-Sit`, `SemRel=syn(onymy)`, and the value of `Abstract Definition`, i.e. `` `Dyn-Sit$_{V1}$ sth is to perform Dyn-Sit$_{N2}$' ``.

L_1	L_2	Schema_{L1}	Schema_{L2}	Concrete Definition
laver	*lavage*	X	Xage	'$\texttt{laver}_{V1}$ sth is to perform $\texttt{lavage}_{N2}$'
classer	*classement*	X	Xment	'$\texttt{classer}_{V1}$ sth is to perform $\texttt{classement}_{N2}$'
planter	*plantation*	X	Xation	'$\texttt{planter}_{V1}$ sth is to perform $\texttt{plantation}_{N2}$'
souder	*soudure*	X	Xure	'$\texttt{souder}_{V1}$ sth is to perform $\texttt{soudure}_{N2}$'
découper	*découpe*	X	X	'$\texttt{découper}_{V1}$ sth is to perform $\texttt{découpe}_{N2}$'

Table 15. Démonette. Affix rivalry

4.3. Meaning-form discrepancies in **Démonette**

We saw how Démonette's set of features can be used to represent almost any type of derivation: regular affixation (*laver→lavage*), conversion, back-formation and affix rivalry. The independence between semantic descriptions (e.g. `SemRel`), morphological structures (e.g. `Morphological Matching`) and structural properties of relations (e.g. `Orientation`) is the key to the descriptive power of this set of features.

With these features, it is also possible to describe the asymmetrical parasynthetic constructions presented in Table 1(b) and in Table 6. The description of these meaning-form discrepancies only requires the addition of two values, `formal-motivation` and `semantic-motivation`, to the feature `Complexity`. Table 16 shows how these values are used[6].

As we discussed above (§ 3.4), parasynthetic formations have distinct formal and semantic motivations. For instance, the forms of *multicellulaire* and *pluricellulaire* are derived from the form of *cellulaire* and their meaning is derived from the meaning of *cellule*. For these parasynthetic forms, the description in Démonette is split into two entries, one for the formal motivation (Table 16(d, e)) and the other for the semantic motivation (Table 16(b, c)). In the first, `Complexity` has the `form(al)-motiv(ation)` value and the semantic features are all left blank. In the other, the semantic relation is `plurality` and the value of `Complexity` is `sem(antic)-motiv(ation)`. This value indicates that the relation is semantically grounded but it is not morphologi-

[6]For the sake of space, the relations are listed in only one direction.

	L_1	L_2	Morph Match	Orientation	Complexity	SemRel
a.	*cellule*	*cellulaire*	X/Xaire	a2d	simple	relation
b.	*cellule*	*multicellulaire*	X/multiXaire	a2d	sem-motiv	plurality
c.	*cellule*	*pluricellulaire*	X/pluriXaire	a2d	sem-motiv	plurality
d.	*cellulaire*	*multicellulaire*	X/multiX	a2d	form-motiv	—
e.	*cellulaire*	*pluricellulaire*	X/pluriX	a2d	form-motiv	—
f.	*pluricellulaire*	*multicellulaire*	pluriX/multiX	indirect	simple	synonymy

Table 16. Démonette. Description of the parasynthetic relations in the family of cellule_N

cally: the values of the feature `Morpho(logical) Match(ing)` (i.e. `X/multiXaire` and `X/pluriXaire`) are not used for regular derivations.

With the values `sem-motiv` and `form-motiv`, Démonette can independently represent relations in the formal and semantic paradigms just as in ParaDis and thus becomes a large-scale formalization of this model: a relation with `Complexity=form-motiv` only belongs to the formal network (no semantic counterpart) while a relation with `Complexity=sem-motiv` only belongs to the semantic network. The other values for `Complexity`, that is, `simple` and `complex`, characterize compositional relations: for instance, the base/derivative regular relation (*cellule*, *cellulaire*) in Table 16(a), and the indirect, prefix replacement relation in the pair of synonyms (*multicellulaire*, *pluricellulaire*) in Table 16(f).

5. Conclusion

In this article, we have presented Démonette and its theoretical background. The resource is under development and many of the results we discussed are still partial. Our goal is to provide a semantically and formally homogeneous description of French derivational morphology, both regular and non-canonical, by combining principles taken from lexeme-based morphology and paradigmatic models of derivation.

We have shown throughout this article that Démonette and ParaDis are actually two sides of the same project. One of the benefits of their joint development is a decisive and mutual enrichment of the two sides. They largely have the same goal which is to model and describe French derivational paradigms. Ultimately, Démonette will provide a playground where all sorts of hypotheses may be tested. Another goal is to provide an effective answer to the question "what does a derivational paradigm look like?". On the other hand, ParaDis addresses the same question from a different angle: "How does paradigmatic derivational morphology work and why do we need it?". The success of this effort owes much to Démonette which helped clarify many ideas morphologists had about derivational paradigms and identify the main principles articulated in ParaDis.

Démonette has a simple, robust and highly redundant representation format where many existing morphological descriptions can be reframed. It is purely relational and only describes the WF processes through the pairs of lexemes they help form. One consequence of the parallel development of Démonette and ParaDis is the importance given to the non-canonical formations in the two sides of the project. Actually, most of the progress brought by this effort comes from the need to have a clean description of the analysis of these formations. It also results in an imbalance in ParaDis where the representational component is fully fledged while the processive one (i.e. the inventory of the constraints that control the filling of the paradigms) remains sketched. Démonette and ParaDis have very similar scopes in terms of phenomena and morphological processes, with one exception: composition. Composition cannot be described in ParaDis because it does not fit in the derivational paradigms defined by the affixations, conversions and all their non-canonical variants. On the other hand, the relations between a compound and its components can easily be represented in Démonette (by means of an additional value `composition` of the feature `Complexity`).

Acknowledgements

This work benefited from the support of the project DEMONEXT ANR-17-CE23-0005 of the French National Research Agency (ANR). We thank the participants in the DEMONEXT project for their contribution to the results presented in this paper.

Bibliography

Amiot, Dany. Plusieurs versus poly-, pluri- et multi-. *Verbum*, 27(4):403–417, 2005.

Anderson, Stephen R. *A-Morphous Morphology*. Cambridge University Press, Cambridge, UK, 1992. doi: 10.1017/CBO9780511586262.

Antoniova, Vesna and Pavol Štekauer. Derivational paradigms within selected conceptual fields – contrastive research. *Facta Universitatis, Series: Linguistics and Literature*, 13(2):61–75, 2015.

Aronoff, Mark. *Word Formation in Generative Grammar*. Linguistic Inquiry Monographs. MIT Press, Cambridge, MA, 1976.

Baayen, R. Harald, Richard Piepenbrock, and Leon Gulikers. The CELEX Lexical Database (Release 2). CD-ROM, 1995. URL `https://catalog.ldc.upenn.edu/LDC96L14`. Linguistic Data Consortium, Philadelphia, PA.

Becker, Thomas. Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In Booij, Geert E. and Jaap van Marle, editors, *Yearbook of Morphology 1993*, pages 1–25. Kluwer Academic Publishers, Dordrecht, 1994. doi: 10.1007/978-94-017-3712-8_1.

Bernhard, Delphine, Bruno Cartoni, and Delphine Tribout. A Task-Based Evaluation of French Morphological Resources and Tools. *Linguistic Issues in Language Technology*, 5(2), 2011. URL `https://halshs.archives-ouvertes.fr/halshs-00746391`.

Blevins, James P. *Word and paradigm morphology*. Oxford University Press, 2016. doi: 10.1093/acprof:oso/9780199593545.001.0001.

Bochner, Harry. *Simplicity in generative morphology*. Mouton de Gruyter, Berlin & New-York, 1993. doi: 10.1515/9783110889307.

Bonami, Olivier and Jana Strnadová. Paradigm structure and predictability in derivational morphology. *Morphology*, 29(2):167–197, 2019. doi: 10.1007/s11525-018-9322-6.

Booij, Geert. *Construction Morphology*. Oxford University Press, Oxford, 2010. doi: 10.1093/acrefore/9780199384655.013.254.

Booij, Geert and Francesca Masini. The role of second order schemas in the construction of complex words. In Bauer, Laurie, Lívia Körtvélyessy, and Pavol Štekauer, editors, *Semantics of complex words*, volume 47, pages 47–66. Springer, Heidelberg, 2015. doi: 10.1007/978-3-319-14102-2_4.

Corbett, Greville G. Canonical derivational morphology. *Word Structure*, 3(2):141–155, 2010. doi: 10.3366/word.2010.0002.

Cotterell, Ryan and Hinrich Schütze. Joint Semantic Synthesis and Morphological Analysis of the Derived Word. *CoRR*, abs/1701.00946, 2017. doi: 10.1162/tacl_a_00003. URL `http://arxiv.org/abs/1701.00946`.

Creutz, Mathias and Krista Lagus. Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora Using Morfessor 1.0. Technical Report A81, Helsinki University of Technology, 2005.

Dahl, Eystein and Antonio Fábregas. Zero Morphemes. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, 2018. doi: 10.1093/acrefore/9780199384655.013.592.

Fellbaum, Christiane, Anne Osherson, and Peter E. Clark. Putting semantics into WordNet's "morphosemantic" links. In *Human Language Technology. Challenges of the Information Society*, volume 5603 of *Lecture Notes in Computer Science Volume*, pages 350–358. Springer, 2009. doi: 10.1007/978-3-642-04235-5_30.

Fillmore, Charles. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280:20 – 32, 12 2006. doi: 10.1111/j.1749-6632.1976.tb25467.x.

Fradin, Bernard. *Nouvelles approches en morphologie*. PUF, Paris, 2003. doi: 10.3917/puf.fradi.2003.01.

Goldsmith, John. Unsupervised Learning of the Morphology of Natural Language. *Computational Linguistics*, 27(2):153–198, 2001. doi: 10.1162/089120101750300490.

Habash, Nizar and Bonnie Dorr. A categorial variation database for English. In *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (NAACL/HLT 2003)*, pages 96–102, Edmonton, 2003. ACL. doi: 10.3115/1073445.1073458.

Hathout, Nabil. Acquisition of morphological families and derivational series from a machine readable dictionary. In Montermini, Fabio, Gilles Boyé, and Jesse Tseng, editors, *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, Somerville, MA, 2009. Cascadilla Proceedings Project. URL `https://arxiv.org/abs/0905.1609`.

Hathout, Nabil. Une approche topologique de la construction des mots : propositions théoriques et application à la préfixation en *anti-*. In *Des unités morphologiques au lexique*, pages 251–318. Hermès Science-Lavoisier, Paris, 2011.

Hathout, Nabil and Fiammetta Namer. Discrepancy between form and meaning in Word Formation: the case of over- and under-marking in French. In Rainer, Franz, Wolfgang U. Dressler, Francesco Gardani, and Hans Christian Luschützky, editors, *Morphology and meaning*, pages 177–190. John Benjamins, Amsterdam, 2014a. doi: 10.1075/cilt.327.12hat.

Hathout, Nabil and Fiammetta Namer. Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology*, 11(5):125–168, 2014b.

Hathout, Nabil and Fiammetta Namer. 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)*, 2016. URL `http://www.lrec-conf.org/proceedings/lrec2016/pdf/279_Paper.pdf`.

Hathout, Nabil and Fiammetta Namer. La parasynthèse à travers les modèles : des RCL au ParaDis. In Bonami, Olivier, Gilles Boyé, Georgette Dal, Hélène Giraudo, and Fiammetta Namer, editors, *The lexeme in descriptive and theroretical morphology*, Empirically Oriented Theoretical Morphology and Syntax, pages 365–399. Language science Press, Berlin, 2018. URL `http://langsci-press.org/catalog/book/165`.

Hockett, Charles Francis. Two models of grammatical description. *Words*, 10:210–234, 1954. doi: 10.1080/00437956.1954.11659524.

Iacobini, Claudio. Parasynthesis in Morphology. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, Oxford, 2020. doi: 10.1093/acrefore/9780199384655.013.509.

Koenig, Jean-Pierre. *Lexical Relations*. CSLI Publications, Stanford, CA, 1999.

Kyjánek, Lukáš. Morphological Resources of Derivational Word-Formation Relations. Technical Report 61, ÚFAL - Charles University, Prague, 2018.

Lafourcade, Mathieu and Alain Joubert. JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes. In *JADT'08 : Journées internationales d'Analyse statistiques des Données Textuelles*, pages 657–666, France, March 2008. URL `https://hal-lirmm.ccsd.cnrs.fr/lirmm-00358848`.

Litta, Eleonora, Marco Passarotti, and Chris Culy. Formatio formosa est. Building a Word Formation Based Lexicon for Latin. In *Proceedings of the Third Italian Conference on Computational Linguistics, CLiC-it 2016*, pages 185–189, Naples, Italy, 2016. URL `http://ceur-ws.org/Vol-1749/paper32.pdf`.

Manova, Stela. Subtraction. In Manova, Stela, editor, *Understanding Morphological Rules - With Special Emphasis on Conversion and Subtraction in Bulgarian, Russian and Serbo-Croatian*, volume 1, pages 125–172. Springer, Dordrecht, 2011. doi: https://doi.org/10.1007/978-90-481-9547-3.

McCarthy, John and Alan Prince. Prosodic Morphology: Constraint Interaction and Satisfaction. Technical report 3, Rutgers University Center for Cognitive Science, New Brunswick, NJ, 1993. URL `https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1013&context=linguist_faculty_pubs`.

Mel'čuk, Igor. Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In Wanner, Leo, editor, *Lexical Functions in Lexicography and Natural Language Processing*, pages 37–102, Amsterdam/Philadelphia, 1996. John Benjamins.

Miller, Georges A., Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4):335–391, 1990. doi: 10.1093/ijl/3.4.235.

Namer, Fiammetta. *Morphologie, lexique et traitement automatique des langues : L'analyseur DériF*. Hermès Science-Lavoisier, Paris, 2009.

Namer, Fiammetta. A Rule-Based Morphosemantic Analyzer for French for a Fine-Grained Semantic Annotation of Texts. In Mahlow, Cerstin and Michael Piotrowski, editors, *SFCM 2013*, CCIS 380, pages 93–115. Springer, Heidelberg, 2013. doi: 10.1007/978-3-642-40486-3_6.

Namer, Fiammetta, Nabil Hathout, and Stéphanie Lignon. Adding morpho-phonology into a French morpho-semantic resource: Demonette. In Litta, Eleonora and Marco Passarotti, editors, *Proceedings of the First Workshop in Resources and Tools for Derivational Morphology (DeriMo),*, pages 49–60, Milano, Italy, oct 2017. EDUCatt. ISBN 978-88-9335-225-3.

Sanacore, Daniele, Nabil Hathout, and Fiammetta Namer. Semantic descriptions of French derivational relations in a families-and-paradigms framework. In Žabokrtský, Zdeněk, Magda Ševčíková, Eleonora Litta, and Marco Passarotti, editors, *Derimo - Second International Workshop on Resources and Tools for Derivational Morphology*, pages 15–24. Charles University, Prague, 2019. ISBN ISBN 978-80-88132-08-0.

Ševčíková, Magda, Adéla Kalužová, and Zdeněk Žabokrtský. Identification of aspectual pairs of verbs derived by suffixation in the lexical database DeriNet. In *Proceedings of the Workshop on Resources and Tools for Derivational Morphology (DeriMo)*, pages 105–116, Milano, Italy, 2017. EDUCatt. ISBN 978-88-9335-225-3.

Shafaei, Elnaz, Diego Frassinelli, Gabriella Lapesa, and Sebastian Padó. DErivCELEX: Development and Evaluation of a German Derivational Morphology Lexicon based on CELEX. In Litta, Eleonora and Marco Passarotti, editors, *Proceedings of the First Workshop in Resources and Tools for Derivational Morphology (DeriMo),*, pages 117–128, Milano, Italy, oct 2017. EDUCatt. ISBN 978-88-9335-225-3.

Šojat, Krešimir, Matea Srebačić, Marko Tadić, and Tin Pavelić. CroDeriV: a new resource for processing Croatian morphology. In Chair), Nicoletta Calzolari (Conference, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.

Spencer, Andrew. *Lexical relatedness*. Oxford University Press, Oxford, 2013. doi: 10.1093/acprof:oso/9780199679928.003.0003.

Steiner, Petra and Josef Ruppenhofer. Building a Morphological Treebank for German from a Linguistic Database. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 3882–3889, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA). URL `https://www.aclweb.org/anthology/L18-1613`.

Štekauer, Pavol. Derivational Paradigms. In Lieber, Rochelle and Pavol Štekauer, editors, *The Oxford Handbook of Derivational Morphology*, pages 354–369. Oxford, Oxford University Press, Oxford, 2014. doi: 10.1093/oxfordhb/9780199641642.013.0020.

Talamo, Luigi, Chiara Celata, and Pier Marco Bertinetto. DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure*, 9(1):72–102, 2016. doi: 10.3366/word.2016.0087.

Thornton, Anna M. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure*, 5(2):183–207, 2012. doi: 10.3366/word.2012.0026.

Todaro, Giuseppina. *Nomi (e aggettivi) che diventano verbi tramite prefissazione: quel che resta della parasintesi*. PhD thesis, Tesi di dottorato, Università Roma Tre et Université Toulouse Jean-Jaurès, 2017.

Tribout, Delphine. Verbal stem space and verb to noun conversion in French. *Word Structure*, 5 (1):109–128, 2012. doi: 10.3366/word.2012.0022.

Vidra, Jonáš, Zdeněk Žabokrtský, Lukáš Kyjánek, Magda Ševčíková, and Šárka Dohnalová. DeriNet 2.0, 2019. URL `http://hdl.handle.net/11234/1-2995`. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Žabokrtský, Zdeněk, Magda Ševčíková, Milan Straka, Jonáš Vidra, and Adéla Limburská. Merging Data Resources for Inflectional and Derivational Morphology in Czech. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1307–1314, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA). URL `https://www.aclweb.org/anthology/L16-1208`.

Zeller, Britta D, Jan Snajder, and Sebastian Padó. DErivBase: Inducing and Evaluating a Derivational Morphology Resource for German. In *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1201–1211, Sofia, Bulgaria, 2013. URL `https://www.aclweb.org/anthology/P13-1118.pdf`.

Address for correspondence:
Fiammetta Namer
`fiammetta.namer@univ-lorraine.fr`
Université de Lorraine, CLSH
UFR SHS, dépt SDL
23 bd Albert 1er, BP 60446
54001 NANCY CEDEX, France