# Démonette-2, a derivational database for French with broad lexical coverage and fine-grained morphological descriptions

*Démonette-2, une base de données dérivationnelle du français à large couverture lexicale munie de descriptions morphologiques détaillées*

Fiammetta Namer

UMR 7118 – ATILF, CNRS & Université de Lorraine

fiammetta.namer@univ-lorraine.fr

Nabil Hathout

UMR 5263 – CLLE CNRS & Université Toulouse-Jean-Jaurès

nabil.hathout@univ-tlse2.fr

Avec : Dany Amiot(3), Lucie Barque(4), Olivier Bonami(4), Gilles Boyé(5), Basilio Calderone(2), Julie Cattini(6), Georgette Dal(3), Alexander Delaporte(7), Guillaume Duboisdindien(3), Achille Falaise(4), Natalia Grabar(3), Pauline Haas(8), Frédérique Henry(1), Mathilde Huguin(9), Nyoman Juniarta(10), Loïc Liégeois(4), Stéphanie Lignon(1), Lucie Macchi(3), Grigoriy Manucharian(11), Caroline Masson(12), Fabio Montermini(2), Nadejda Okinina(13), Franck Sajous(2), Daniele Sanacore(2), Mai Thi Tran(3), Juliette Thuilier(2), Yannick Toussaint(14), Delphine Tribout(3)

(1)     UMR 7118 – ATILF, CNRS & Université de Lorraine

(2)     UMR 5263 – CLLE CNRS & Université Toulouse Jean-Jaurès

(3)     UMR 8163 – STL CNRS & Université de Lille

(4)     UMR 7110 – LLF CNRS & Université Paris-Cité

(5)     UMR 5263 – CLLE CNRS & Université Bordeaux

(6)     Orthophoniste, Luxembourg

(7)     UMR8563 – CRLAO CNRS & EHESS & INALCO

(8)     UMR 8094 - Lattice CNRS, ENS & Université Paris Sorbonne Nouvelle

(9)     INIST CNRS

(10)    CapGemini, Nice

(11)    UMR 7114 – MoDyCo CNRS & Université de Nanterre

(12)    EA7345 - CLESTHIA Université Paris Sorbonne Nouvelle

(13)    SONOS Vox

(14)    UMR 7503 – LORIA CNRS, Université de Lorraine & Inria

## Abstract

Morphological databases play an important role in linguistic research today. While several exist for the study of inflectional morphology in French, there is still a lack of resources for derivational morphology. This article presents Démonette-2, a new release of the French derivational morphology database Démonette. Developed in the context of the Démonext project, this database provides a framework in which it is possible to integrate various existing derivational resources. Démonette is characterized by its relational nature, which makes it possible to describe an extensive set of word formation patterns, including suffixation, prefixation, and a variety of non-canonical derivational processes such as conversion and parasynthetic formations. It reconciles broad coverage and fine-grained descriptions and is suitable for different audiences: morphologists working in different theoretical frameworks, teachers, speech therapists, NLP specialists, etc. The article presents in detail the structure and content of Démonette, its evolution since the first version, and its query interface.

**Keywords:** derivational morphology, paradigm-based morphology, derivational families, database, feeding from existing resources, form-meaning discrepancy, web interface.

## Résumé

Les bases de données morphologiques jouent aujourd'hui un rôle important dans les recherches en linguistique. S'il en existe plusieurs pour les études de morphologie flexionnelle pour le français, l'offre reste insuffisante pour la morphologie dérivationnelle. Cet article présente Démonette-2, une nouvelle version de la base de données morphologique dérivationnelle du français Démonette. Développée dans le cadre du projet Démonext, cette base offre un cadre dans lequel il est possible d'intégrer diverses ressources dérivationnelles existantes. Démonette se caractérise par sa nature relationnelle qui permet d'y décrire un ensemble étendu de constructions morphologiques qui inclut les procédés de suffixation, de préfixation, et une grande variété de dérivations non canoniques comme la conversion et les constructions parasynthétiques. Elle concilie couverture étendue et finesse des descriptions et est adaptée à différents publics cibles : morphologues travaillant dans différents cadres théoriques, enseignants, orthophonistes, spécialistes de TAL, etc. L'article présente en détail la structure et le contenu de Démonette, son évolution depuis la première version et son interface d'interrogation.

**Mots-clefs :** morphologie dérivationnelle, morphologie paradigmatique, familles dérivationnelles, alimentation par des ressources existantes, décalage sens-forme, interface web.

# 1. Introduction

Morphological databases play an increasingly important role in today's linguistic research. Several exist for French, such as Lexique (New, 2006) for lexical studies and inflectional morphology. The situation is quite different for derivational morphology resources, which have long been few, modest in size, under copyright, heterogeneous in format and content, difficult to use, etc. This article describes Démonette-2, a new release of the French derivational morphology database Démonette, developed within the Demonext project, which provides solutions to some of the above-mentioned problems. Démonette was designed as a framework for integrating various existing derivational resources, harmonizing their format, and compiling their entries into a single database.

A notable feature of Démonette is its relational nature: its entries describe relations between morphologically related pairs of lexemes and are not limited to the analysis of derivatives with respect to their base. For example, it contains entries for the relation between *chanteur*.Nm 'male singer' and *chanter*.V 'to sing', between *danseur*.Nm 'male dancer' and *danseuse*.Nf 'female dancer', and between *sonore*.A 'sonorous' and *insonoriser*.V 'to soundproof'. As a result, the descriptions in Démonette are in line with the relational approach to derivation (Jackendoff & Audring, 2020), are consistent with the principles of lexematic morphology (Aronoff, 1994; Booij, 2010; Fradin, 2003; Matthews, 1972, 1991), and meet the needs of research in paradigmatic derivational morphology (Robins, 1959; Bauer, 1997; Booij, 2008; Booij & Masini, 2015; Bonami & Strnadová, 2019; Fradin, 2018; Hathout & Namer, 2019; 2022). Note that currently Démonette only describes morphological relations in synchrony.

Version 2 is a major evolution of Démonette. Its development was guided by several objectives, one of which being to create a resource that combines broad coverage and fine-grained descriptions and is adapted to different target audiences: morphologists working in different theoretical frameworks, teachers, speech therapists, natural language processing (NLP) specialists, etc. Another objective is to preserve as much as possible the analyses of the resources it integrates.

The cumulative integration of resources also allows Démonette to benefit from their quality, since most of them have been produced in the context of studies of different processes and morphological phenomena. As a result, it covers a significant part of the French lexicon while describing an important number of remarkable morphological processes such as conversion and paradigmatic phenomena (e.g., parasynthetic constructions) that resist the canonical binary and oriented description between a derived word and its base. It includes both central derivational word formations (WF), such as suffixation in *-age* used to form action nouns (*nettoyage*.Nm 'cleaning'), or suffixation in *-eur* used to form agent nouns (*contrôleur*.Nm 'inspector'), and more marked and less often described formations, such as suffixation in *-at* (*matriarcat*.Nm 'matriarchy'), discussed in Roché and Plénat (2012) among others, or in *-itude* (*exactitude*.Nf 'accuracy'), see Koehl and Lignon (2014).

## 2. Related work

There are several databases that can be used for lexical and inflectional morphology studies in French. The best known is Lexique[1] (New, 2006), designed for psycholinguistic experiments. Lexique provides an extensive set of morphological, phonological and distributional features that allow a careful selection of experimental material. Another original quality of Lexique is that it includes words (i.e., inflected forms) from an "authentic" corpus made up of subtitles from 9 474 movies, totaling some 50 million occurrences. On the other hand, its size is relatively small (142 694 entries in Lexique-3.83) compared to inflectional lexicons such as Morphalou[2] (Romary et al., 2004), whose version 1.0.1 contains 524 725 entries, Flexique[3] (Bonami et al., 2014), which has 363 293 inflected forms, or GLàFF[4] (Sajous et al., 2013; Hathout et al., 2014), which has more than 1.4 million entries.

In contrast, there are few resources for derivational morphology in French. One of the first is Démonette-1 (Hathout & Namer, 2014a, 2014c). More recently, a French derivational lexicon has been integrated into the UniMorph[5] database (Batsuren et al., 2022). This lexicon describes 73 256 morphologically related lexeme pairs. UniMorph is a resource that provides inflectional and derivational morphological descriptions for 168 languages. It is mainly used in NLP, especially for SIGMORPHON tasks that need large datasets to train models and for which the amount of data is a critical factor. On the other hand, morphological description is minimal. UniMorph provides only the morpheme corresponding to the last operation in the derivational history of the derived lexemes, which does not allow direct identification of some non-canonical WFs such as the parasynthetic formations.

For English, the supply of derivational resources is more important. One of the "historical" databases for this language is CELEX[6] (Baayen et al., 1995), which was one of the first to provide the detailed and complete derivational descriptions needed to select experimental material, especially for use in psycholinguistics. CELEX is a resource produced by lexicographers and is therefore perfectly homogeneous. On the other hand, it is relatively small. The English derivational database provides morphological analyses of 45 968 entries of derived lexemes (affixed or converted words, compounds, etc.) and contains 8 490 entries of simple lexemes. For comparison, the English section of UniMorph provides derivational descriptions for 225 131 lexeme pairs.

In recent years, other derivational resources have been developed for other languages, such as DeriNet for Czech (Vidra et al., 2019). Kyjánek (2018) provides a comprehensive overview of these

---

[1]    http://www.lexique.org/

[2]    https://www.ortolang.fr/market/lexicons/morphalou

[3]    http://www.llf.cnrs.fr/fr/flexique-fr.php

[4]    http://redac.univ-tlse2.fr/lexicons/glaff_en.html

[5]    https://unimorph.github.io/

[6]    CELEX also provides derivational descriptions for Dutch and German. It is available at https://catalog.ldc.upenn.edu/LDC96L14

resources and their distinctive features. Finally, we should mention the recent creation of Universal Derivation, a multilingual derivational database built from existing resources (Kyjánek et al., 2020; Kyjánek et al., 2021; Žabokrtský et al., 2022). The goals of Universal Derivation are like those of Démonette. However, the two databases differ in that the derivational relations in Universal Derivation form rooted trees and therefore only connect derived words to their bases.

## 3. Constants and evolutions of Démonette

The organization and distribution of information in Démonette is based on a set of principles that remain stable from one version of the database to the next (Section 3.1). Besides these constants, Démonette has undergone continuous improvements and additions, which we discuss in section 3.2. As mentioned above, version 2 is a major redesign of the database, and the number of entries has increased significantly, from 31 204 in Démonette-1.0 to 271 698 in Démonette-2.

### 3.1 Constants

### 3.1.1 General philosophy

The fundamental principle of Démonette is that its entries are relations that connect two words of the same derivational family: one connects *laver*.V 'to wash' to *lavage*.Nm 'washing', another, *lavage*.Nm to *lavable*.A 'washable', another, *laver*.V to *inlavable*.A 'unwashable', and so on. A second principle is that entries are described as a flat structure, regardless of the morphological complexity of the relation, and of the distance that separate the two lexemes in their derivational family graph. For example, the entry *laver*.V-*inlavable*.A does not include a complete description of the internal structure of the adjective, which would be represented in parenthetical form as [in[[laver] able]] in the morpheme tradition; the description only states that the second lexeme has a prefix (*in-*) and a suffix (-*able*) that the first one does not have. Since the relations *laver*.V-*lavable*.A and *lavable*.A-*inlavable*.A are encoded in Démonette, we can combine the properties of these two relations to reconstruct the derivational path between *laver*.V and *inlavable*.A, and consequently the morphological structure of *inlavable*.A relative to *laver*.V. The other properties of the database presented below are: the relational conception of morphology (Section 3.1.2); the diversity of processes represented (Section 3.1.3); the redundancy of descriptions (Section 3.1.4); the symmetry of relations (Section 3.1.6); the "ecumenism" in the derivational approach claimed in the coding of properties (Section 3.1.7). Another principle at the heart of Démonette's conception, mentioned in Section 1, is that derivation is considered within a lexematic approach to morphology. According to this approach, morphology is relational, the morpheme is "dereified", units are described on three levels (formal, categorical, semantic), and these levels are exploited by morphological relations simultaneously and independently. Finally, each pair is analyzed "locally", independently of the rest of the lexicon.

### 3.1.2 A relational conception of derivational morphology

A distinctive feature of Démonette is that it describes morphological relations and not morphologically complex lexemes (Hathout & Namer, 2014a, 2014c) as it is the case in other databases such as Lexique or UniMorph whose entries are inflected forms or derived lexemes. It fits into a relational approach to derivational morphology (Jackendoff & Audring, 2020) and a paradigmatic approach (Robins, 1959; Bauer, 1997; Booij, 2008; Booij & Masini, 2015; Bonami & Strnadová, 2019; Fradin, 2018; Hathout & Namer, 2019; 2022; see Štekauer, 2014 for an overview of the issue). In these approaches, lexeme properties are partially determined by the relations they are part of. For example, the noun *militantisme*.Nm 'militancy' is both a lexeme in *-antisme* when considered in its relation to the verb *militer*.V 'to be an activist' and a lexeme in *-isme* when considered in its relation to the noun *militant*.Nm 'activist' (Section 3.1.1). The same goes for its semantic and formal properties: on the semantic level, the relation between *militantisme*.Nm and *militer*.V ('doctrine related to the act of militating') adds to the one that links *militantisme*.Nm to *militant*.Nm ('doctrine of the activists').

### 3.1.3 Diversity of relations

Démonette's entries can describe a wide range of morphological constructions, including suffixation, prefixation, and most non-canonical derivations like conversion (Tribout, 2012) and parasynthetic formations (Hathout & Namer 2014b, 2018; Iacobini, 2020). More generally, Démonette can account for all types of binary derivational relations as between a noun such as *entoilage*.Nm 'interfacing' and its base *entoiler*.V 'to stiffens with canvas', or one of the ancestors of its base, i.e., *toile*.Nf 'canvas', between two lexemes derived from the same base, such as *entoilage*.Nm 'interfacing' and *entoilement*.Nm 'interfacing', or between two more distant members of the same family, like *toiliste*.Nm 'worker who stiffens with canvas' and *entoilage*.Nm.

Ordinary compounding is the only process excluded because it involves three lexemes: one compound and two components (e.g., *porte-fenêtre*.Nf 'French window', *porte*.Nf 'door', *fenêtre*.Nf 'window'). Conversely, neoclassical compounding (e.g., *biodégradable*.A 'biodegradable') can be conceived as a binary "base → derivative" relation like affixal derivations and therefore included in Démonette. Reasons for this include the fixed position of the components (e.g., *-logue* is always placed on the right in the neoclassical compound, while *bio-* is always placed on the left), the deconceptualization of their content (e.g., *-logue* has lost the sense of "speech" that its Greek ancestor *logos* had, in favor of a fuzzy content evoking the notion of "specialist", see Namer & Villoing, 2014), and their use in numerous and recent creations by speakers of French who don't necessarily have any knowledge of Latin or Greek (e. g., *je suis hotélophobe mais vacançophile* 'I'm a hotelophobe but a vacationophile'). See Lasserre and Montermini (2014) for further arguments in favor of the grammaticalisation of neoclassical components. Note that standard compounding could be described as two binary relations linking a compound to its two components. However, this description would be incomplete in that in each of these relations the other component would

become a kind of exponent. Therefore, such a description would not account for the real contribution of the components to the compound word.

### 3.1.4 Redundancy

Another important feature of Démonette is the redundancy of the descriptions. Some information is duplicated in several entries. For example, the fact that the lexeme *laver*.V 'to wash' is a verb is described in all entries in which it appears. `Démonette` may also contain multiple descriptions of the same relation, for example when they come from different sources. When the authors of two resources each propose an analysis of the same pair of lexemes, the two analyses are described in two database entries, distinguished by the identifier of the original source of the analysis. In this way, the description of any entry in the database is independent of the description of any other entries so that we can add or delete entries without the risk of making the others incomplete or inconsistent. On the other hand, the duplication of information can lead to inconsistencies between different descriptions of the same information. For example, the pair *candisation.*Nf-*candir.*V 'candisation'-'to candy' is analyzed as formed by a single, direct *-isation* suffixation linking a descendant to an ascendant, but this same relation is analyzed as an indirect one in other pairs such as *chromisation.*Nf-*chromer.*V 'chromatization'-'to chromate'.

### 3.1.5 Sourcing

Démonette is populated by several resources that have undergone manual cleaning, standardization of their format, and possibly revision of some of their descriptions. Table 1 summarizes the contribution of these resources to Démonette-2.

| Resource | Authors | Description | Number of entries |
|---|---|---|---|
| **Convers** | Tribout (2010) | noun/verb and verb/noun pairs in a relation of conversion | 3 241 |
| **Démonette-1** | Hathout and Namer (2014) | derivational relations between verbs, action nouns, agent nouns and agentive adjectives | 96 027 |
| **DiMoC** | Roché (2004, 2005, 2011a,b, 2017b), Roché and Plénat (2012) | collection of derivational morphological databases annotated with a large set of semantic, morphophonological and diachronic features | 46 211 |
| **Denom** | Strnadová (2014) | database of denominal adjectives that records a variety of properties, including derivational, phonological, distributional, etc. | 15 094 |
| **DériF** | Namer (2009) | output of the DériF morphological analyzer | 9 435 |
| **Mordan** | Koehl (2012) | database of quality nouns derived from adjectives | 4 434 |

Table 1. Sources of the descriptions contained in Démonette-2

Démonette's entries indicate the origin of each piece of information they contain. The sourcing allows users to select the descriptions they are interested in according to the resources they originate from.

### 3.1.6 Relations are symmetric

In Démonette, the morphological relations are symmetric. Symmetry is natural for indirect relations, e.g., between two lexemes derived from the same base (Section 4.2.2), as in the case of the pairs *arroseur*.Nm-*arroseuse*.Nf 'male waterer', 'female waterer' and *arroseuse*.Nf-*arroseur*.Nm[7], because in both pairs, the first lexeme is the correspondent of the second one. Symmetry is extended to ascending and descending relations (i.e., between a base and one of its derived words, and between a derived word and its base), since it is generally possible to define each of the two lexemes with respect to the other. For example, in the relation *arroser*.V-*arroseur*.Nm 'to water'-'male waterer', 'male waterer' is "the person who waters" and 'to water' could be paraphrased as "to do what a waterer does".

---

[7]    In Démonette, male and female correspondents of an animate entity are considered to be different lexemes.

### 3.1.7 A pan-theoretical approach to morphology

Démonette provides a "flat" description of the information it contains. This format makes it easy to feed the database with new resources, while remaining as faithful as possible to the analyses they contain. Its descriptions conform to the lexematic approaches to morphology, but can be reformulated in a way that makes them consistent with other theoretical frameworks, such as the morpheme-based or rule-based approaches (Hockett, 1954). For example, one can easily reconstruct morpheme representations of derived lexemes from the direct ascending relations in their derivational history.

### 3.1.8 Independence of morphological, formal, categorical and semantic information

Remember that in Démonette, the description of a derivational relation is inspired by the principles of lexeme-based morphology (Section 3.1.1). The records in the database consist therefore of four types of information, formal (or phonological), categorical, semantic and morphological, represented independently in four groups of fields in the tables. As a result, Démonette is consistent with the ParaDis model (Namer & Hathout, 2020; Hathout & Namer, 2022).

## 3.2 Developments since the first version of Démonette

The main changes in the first versions of Démonette fall in three broad categories:

1.  Breaking down the features into more basic properties in order to separate the different information in the descriptions as much as possible.

2.  Refining feature values to make them more explicit and better able to describe a wide range of derivational relations.

3.  Increasing the number of entries by adding new resources to the database.

**Démonette-1.0.** The first version of Démonette is presented in (Hathout & Namer, 2014a, 2014c) but has not been published. It consists of 32 600 entries composed of indirect relations between lexemes from the Morphonette derivational lexicon (Hathout, 2011a) and from base to derivative relations resulting from the analysis of TLFnome[8] entries by the DériF derivational analyzer (Namer, 2009, 2013). This first version focuses mainly on the description of meaning. The lexeme pairs are provided with semantic types, relational definitions of their meaning (with respect to the other lexeme in the pair), and semantic patterns, i.e., abstract representations of these definitions in which the type of the other lexeme is substituted for its form. This version differs from the following ones in several ways:

• the indication of the source is global (Dérif or Morphonette). It is given once for the entry and all its features (whereas in Démonette-2 the source is specified separately for each feature);

• word form and category are combined into a single feature (e.g., `compilation=N` for the noun *compilation.*Nf 'compilation');

---

8    TLFnome contains the word list of the *Trésor de la Langue Française* dictionary.

- the entries do not provide WF schemes;
- WF is limited to seven verb-based suffixations: *-age, -ion, -ment, -eur, -euse, -rice, -if*;
- the entries have direct and opposite definitions. The direct definitions describe the meaning of the first lexeme with respect to the second, and the opposite definitions describe the meaning of the second lexeme with respect to the first.

**Démonette-1.1** is the first version of the database to be released. This version contains 77 323 entries and brings several improvements compared to version 1.0:
- the description of word forms and categories are in separated fields;
- the entries include the origin of each piece of information;
- the type and exponent (i.e., suffixes) of the WF are given for the two lexemes;
- the complexity and orientation of the relations are provided;
- the entries describe only the meaning of the first lexeme in relation to the second;
- the number and variety of WF processes are increased;
- semantic types `@ACT` (action) and `@RES` (result) are distinguished;
- entries describing direct relations contain a graphemic representation of the stem of the derived word;
- entries that come from more than one resource are duplicated.

**Demonette-1.2** has also been released. It contains essentially the same information as version 1.1, but differs from the latter in that it has been supplemented with entries from the VerbAction database. Démonette-1.2 contains 96 072 entries, 25 additional deverbal action noun exponents (*-ure, -ance*, etc.) and 1 540 entries that describe conversions.

**Demonette-1.3** was presented in Namer et al. (2017) but was not published. The main difference with version 1.2 is the inclusion of 71 340 entries from the Lexeur lexicon (Wauquier et al., 2020), which provides partial morphological families[9] of 5 974 French agent nouns ending in *-eur*. In addition, Démonette-1.3 adds phonemic representations from the GLàFF lexicon to the entries and implements a pilot study aimed at automatically computing phonological variations between the lexemes in each pair. In total, the database contains 167 367 entries. Pairs present in more than one resource are described by different entries.

## 4. Description of derivational morphology in Démonette-2

### 4.1 Splitting the descriptions into several tables

An important difference between Démonette-2 and Démonette-1 is the distribution of the descriptions in three co-indexed tables: the table of relations, the table of lexemes and the table of

---

[9]    Following (Hathout, 2009, 2011b), we define a morphological family (also called derivational family or word family) as a set of lexemes connected by derivational relations.

families. The table of relations contains the properties of derivational relations, while the table of lexemes describes the information that is specific to the lexemes, regardless of their possible relations with other members of their families. This separation avoids some duplication of the phonological and semantic features of the lexemes.

•   The phonological description of derivational relations, introduced in Démonette-1.3 includes a phonological analysis of the stem variation between the lexemes of the entries. We consider a variation to exist when the lexemes do not have a common stem as in *deux*.Num-*double*.A 'two'- 'double' (/dø/-/dubl/) where /du/ is not a stem of the numeral *deux*. The identification of variations is based on the comparison of the phonemic transcriptions of the inflectional paradigms of the two lexemes. The distribution of information in several tables makes it possible to record the inflectional paradigms of the lexemes on which the phonological analysis is based only once.

•   The morphosemantic description of its entries distinguishes Démonette from other morphological databases since its first version. Démonette describes the semantic relations between pairs of lexemes and the semantic properties of the lexemes that are relevant for this relation (Table 2). The inclusion of these features was possible in the first versions of Démonette, because its relations were all centered around a verbal predicate. They fall within a typical action network (see Fradin, 2020, 2021; Roché 2017a, and Roché's contribution, in the same volume).

| W1 | W2 | Semantic type of W1 | Semantic type of W2 | Definition of W2 |
|---|---|---|---|---|
| élire.V 'to elect' | électeur.Nm 'elector' | @ | @AGM | (male agent OR instrument) of élire |
| aspirer.V 'to vacum' | aspirateur.Nm 'vacum cleaner' | @ | @AGM | (male agent OR instrument) of aspirer |

Table 2. Example of the morpho-semantic features in Démonette-1

In these descriptions, lexemes only have one type. This simple single type descriptions could not be maintained in Démonette-2 because it includes more diverse derivational relations. The solution adopted in Démonette-2 separates the semantic properties that depend on the morphological relation from the ones that depend on the ontological nature of the lexemes, independently of their participation in the relation. This is in line with the classical distinction between function and category found in syntax.

Technically, the structure adopted is that of a relational database consisting of three tables whose contents are linked by three sets of keys: the table of lexemes collects the lexical information; the table of relations documents the morphological relations between the lexemes; the table of families describes the word families. Word families are represented as lists of lexeme identifiers. The properties encoded in the other two tables are described in detail in Sections 4.2 and 4.3.

## 4.2 Table of relations

The table of relations describes how lexemes are related to each other in morphological families. Its entries are pairs of lexemes. These pairs correspond to edges in the family graph (Figure 1). The structure of the table follows the same principles as in Démonette-1: the annotations are divided into three independent sets of features, corresponding to the three levels found in the lexematic approaches of morphology. In Démonette-2, these sets of features contain:

1. the identity of the two lexemes W1 and W2: written forms; grammatical categories; identifiers of the corresponding entries in the table of lexemes;
2. the morphological properties of the relation between W1 and W2;
3. the semantic properties of the relation. These properties are not implemented yet.

In addition, entries are identified by unique indices (relation identifiers; RID). Moreover, all relations in Démonette are symmetrical: if the table of relations includes an entry W1-W2, then it also includes an entry W2-W1.[10]

### 4.2.1 Multiple descriptions

One same W1-W2 pair can be analyzed in several ways in Démonette. These analyses are described in separate entries in the database. In the example below, the pair *collectionneur*.Nm-*collectionariat*.Nm 'collector'-'state of being a collector' is directly related (base → derived word) in the entry r120059, where X represents both the form *collectionneur* and the allomorphic stem *collectionnar* of the derived word *collectionariat*.Nm The same pair *collectionneur*.Nm-*collectionariat*.Nm is also indirectly related in the entry r120060. In the second relation, both nouns derive from the verb *collectionner*.V 'to collect' and X represents any one of the stems of the verb.

| RID | W1 | W2 | cstr_1 | cstr_2 |
|---|---|---|---|---|
| r120059 | collectionneur 'collector' | collectionariat 'state of being a collector' | X | Xat |
| r120060 | collectionneur | collectionariat | Xeur | Xariat |

Table 3. The lexeme *collectionneur*.Nm 'collector' appears in two relations in the table of lexemes

A lexeme can therefore be related to more than one member of its family. When this happens, some of its (semantic and morphological) properties are determined by these relations.

### 4.2.2 Description of the derivational relations

Démonette uses six features to describes morphological (i.e., WF) relations: CSTR_1, CSTR_2, TYPE_CSTR_1, TYPE_CSTR_2, COMPLEXITE, ORIENTATION (Table 4).

---

[10]   Rows 1 and 2 in Table 4 present symmetrical entries.

| W1.Cat | W2.Cat | type_cstr_1 | type_cstr_2 | cstr_1 | cstr_2 | complexite | orientation |
|--------|--------|-------------|-------------|--------|--------|------------|-------------|
| lavage.Nm 'washing' | laver.V 'to wash' | suf | NA | Xage | X | simple | des2as |
| laver.V 'to wash' | lavage.Nm 'washing' | NA | suf | X | Xage | simple | as2des |
| relaver.V 'to wash again' | laver.Nm 'to wash' | pre | NA | reX | X | simple | des2as |
| action.Nf 'action' | actionner.V 'to operate' | NA | conv | X | X | simple | as2des |
| accès.Nm 'access' | accéder.V 'to access' | conv | NA | X | X | simple | des2as |
| vol.Nm 'flight' | voler.V 'to fly' | conv | conv | X | X | simple | NA |
| lavage.Nm 'washing' | laveur.N 'washer' | suf | suf | Xage | Xeur | simple | indirect |
| bellicisme.Nm 'warmongering' | belliciste.Nm 'warmonger' | suf | suf | Xisme | Xiste | simple | indirect |
| biodégradable.A 'biodegradable' | dégradable.A 'degradable' | comp | NA | bioX | X | simple | des2as |

Table 4. Examples of pairs where COMPLEXITE = simple in the table of relations

The morphological properties of entries in Démonette are first described using the CSTR_1 and CSTR_2 features. These features describe the possible derivational exponents (i.e., affixes) of W1 and W2. Their values are morphological patterns consisting of a variable X representing the stem common to both lexemes and the derivational exponents of the two lexemes. For example, *lavage* (row 1 in Table 4) is described by an Xage pattern where X represents the stem /lav/ and *-age* is the exponent of the WF. The pattern of the verb *laver* is X (not Xer) because the suffix *-er* is not derivational. This is the mark of the infinitive. Note that some variations are omitted in the patterns, notably when the lemmas in W1 and W2 include different inflectional stems of the same lexeme. For example, the variation between the stems in *déceler*.V-*décèlement*.Nm 'to detect'-'detection' is not described insofar as /desɛl/ is a inflectional stem of the verb *déceler*.V.

CSTR_1 and CSTR_2 are complemented by two other features, TYPE_CSTR_1 and TYPE_CSTR_2, which indicate the type of construction. The value of TYPE_CSTR_1 is suf when W1 is formed by suffixation from W2 (*lavage*.Nm-*laver*.V), pre if W1 is formed by prefixation from W2 (*relavage*.V-*laver*.V), pre-suf if the formation of W1 from W2 involves both prefixation and suffixation (*relavage*.Nm-*laver*.V 'to rewash'-'to wash'), conv if W1 results from a conversion of W2 (*actionner*.V-*action*.Nf 'to

activate'-'action'), `comp` if W1 is a neoclassical compound constructed from W2 (*biodégradable*.A-*dégradable*.A). TYPE_CSTR_2 values are determined in the same way.

The values of the features COMPLEXITE and ORIENTATION characterize the derivational "proximity" of W1 and W2. Relations involving only one derivational operation are considered to be simple (COMPLEXITE = `simple`). These relations can be oriented from an ascendant lexeme to a descendant lexeme (`as2des`) as in *laver*.V-*lavage*.Nm, or in the opposite direction, from a descendant lexeme to an ascendant lexeme (`des2as`) as in *lavage*.Nm-*laver*.V. Conversions are also considered to be simple WFs. However, their orientation cannot always be determined (as shown by Tribout 2010, 2012, among others). When it can be, conversions can be as2des as for *action*.Nf-*actionner*.V or `des2as` as for *accès*.Nm-*accéder*.V 'entrance'- 'to access'. When it is not, the orientation value is NA as for *vol*.Nm-*voler*.V 'theft'-'to thief'. In as2des and des2as oriented conversions, the type of construction of the ascendant lexeme is NA. When orientation is undefined, W1 is annotated as converted from W2 and vice versa.

On the other hand, relations between direct descendants of the same lexeme are considered to be simple, as in the case of *laveur*.Nm-*lavage*.Nm 'washer'-'washing'. The orientation of the pair is then indirect, and the types of construction of the two lexemes correspond to their relations with their common base.

Similarly, neoclassical composition (*biodégradable*.A-*degradable*.A 'biodégradable'-'degradable') is also considered to be simple and direct, following Lassalle & Montermini (2004), who propose to analyze these constructions as affixations. As mentioned (Section 3.1.3), the arguments in favor of such an analysis are: (i) that the position of the components is fixed in the compounds; (ii) that the content of the components is deconceptualized; (iii) that no knowledge of the original language of the neoclassical components is needed to use them.

The relation between W1 and W2 is considered to be simple and indirect if it connects two words derived from the same base (*lavage*.Nm-*laveur*.Nm, 'washing'-'washer') without either of them being derived from the other. Pairs where the common base is not, or no longer, attested in synchrony (*bellicisme*.Nm-*belliciste*.Nm 'warmongering'-'warmonger') are also considered to be simple and indirect. Their relations correspond to what Becker (1993) calls cross-formations (see also Hathout & Namer, 2014b), also called substitutive formations (see among others Bonami & Guzmán, 2023), or second-order constructions (Booij & Masini, 2015).

| W1.Cat | W2.Cat | type_cstr_1 | type_cstr_2 | cstr_1 | cstr_2 | complexite | orientation |
|--------|--------|-------------|-------------|--------|--------|------------|-------------|
| militantisme.Nm 'militancy' | militer.V 'to be an activist' | suf | NA | Xantisme | X | complexe | des2as |
| entoilage.Nm 'interfacing' | toiliste. Adj 'worker who stiffens with canvas' | pre-suf | suf | enXage | Xiste | complexe | indirect |

Table 5. Examples of pairs where COMPLEXITE = complexe in the table of relations

Table 5 shows two examples of complex relations (COMPLEXITE = complexe) in which the derivational formation of W1 with respect to W2 involves at least two steps. In the first, *militantisme*.Nm-*militer*.V 'militancy'-'to be an activist', two successive suffixes are needed to build *militantisme*.Nm from *militer*.V: a suffixation in -*ant* followed by a suffixation in -*isme*. Its analysis involves an intermediate step *militant*.Nm 'activist'. The relation is direct because *militantisme*.Nm is a descendant of *militer*.V (ORIENTATION = des2as). The relation between *entoilage*.Nm 'interfacing' and *toiliste*.Nm 'worker who stiffens with canvas' is also complex, since *toiliste*.Nm is derived from *toile*.Nf 'canvas' by suffixation with -*iste*, while *entoilage*.Nm is derived from *toile*.Nf by prefixation with *en*- and suffixation with -*age*. Three elementary WF operations are therefore necessary to describe the relation between *entoilage*.Nm and *toile*.Nf. Moreover, the relation is indirect because neither of *entoilage*.Nm and *toiliste*.Nm is the ascendant of the other.

| W1.Cat | W2.Cat | type_cstr_1 | type_cstr_2 | cstr_1 | cstr_2 | complexite | orientation |
|---|---|---|---|---|---|---|---|
| mentir.V 'to lie' | mensonge.Nm 'lie' | NA | suf | X | Xonge | accidentel | as2des |
| scolariser.V 'to school' | scolaire.A 'related to school' | suf | NA | Xaire | X | motiv-form | des2as |
| monarchique.A 'monarchic' | anti-monarchique.A 'anti-monarchic' | NA | suf | X | Xique | motiv-form | as2des |
| monarchie.Nf 'monarchy' | anti-monarchique.A 'anti-monarchic' | NA | pre-suf | X | anti-Xique | motiv-sem | as2des |
| hippisme.Nm 'equestrianism' | cheval.Nm 'horse' | suf | NA | Xisme | X | motiv-sem | des2as |
| thérapeutique.A 'therapeutic' | thérapique.A 'therapeutic' | suf | suf | Xeutique | Xique | motiv-sem | indirect |
| *bellicisme*.Nm 'warmongering' | *bellliciste*.Nm 'warmonger' | suf | suf | Xisme | Xiste | simple | indirect |
| *hydroplaner*.V 'to hydroplane' | *hydroplanage*.Nm 'hydroplaning' | NA | suf | X | Xage | simple | des2as |

Table 6. Examples of irregular WF from the table of relations

Table 6 presents several WF often regarded as irregular, and illustrates how the feature combinations in Démonette reflect this. First, the relation *mentir*.V-*mensonge*.Nm is analyzed as accidental (COMPLEXITE = accidentel) because *mentir*.V and *mensonge*.Nm are morphologically related, but the relation cannot be analyzed in synchrony: (i) *-onge* is not an available suffix in French and (ii) /mãs/ is not a stem of the verb *mentir*.

The feature COMPLEXITE is also used to describe relations with a discrepancies between form and meaning, illustrated in Table 6 with the examples of *scolariser*.V and *anti-monarchique*.A (the latter being also know as parasynthetic word formation). These discrepancies are called "extended exponence" by Matthews (1972, p. 82) and "one-to-many correspondence" by Booij (1986); they have also been studied by Stump (2017, p. 69), who speaks of "a one-to-many relation of content to form in the morphology of a single word", and by Hathout and Namer (2014b), who proposed a typology.

In Démonette, discrepancies are analysed by a combination of two relations involving the same complex word W:

- the first one is said to be motivated formally but not semantically (COMPLEXITE = motiv-form) when the form of W is coined on that of the simpler word, but the meaning of W cannot be deduced from that of the other word. For example, the form of *scolariser*.V 'to school' is built on that of *scolaire*.A 'related to school' by suffixation in *-iser*

(modulo the /ɛ/-/a/ variation in the last syllable of the base stem), but the verb does not mean 'to make school-related'. The relation *scolariser*.V-*scolaire*.A is therefore annotated as: CSTR_1 = Xaire, CSTR_2 = X, ORIENTATION = des2as, COMPLEXITE = motiv-form.

- the second relation is said to be motivated semantically but not formally (COMPLEXITE = motiv-sem) when the meaning of W can be derived from that of the simpler word, but the form of W is not directly coined on that of this simpler word. For example, *scolariser*.V can be directly defined from *école*.Nf 'school' (*scolariser*.V 'to school' a child is 'to send the child to school') but there is no direct formal relation between the two lexemes. The relation *scolariser*.V-*école*.Nf is described by the features: CSTR_1 = Xariser, CSTR_2 = X, ORIENTATION = des2as, COMPLEXITE = motiv-sem.

Then, the entry *hippisme*.Nm-*cheval*.Nm in Table 6 illustrates another case of irregular relation. Here, COMPLEXITE = motiv-sem indicates that the stem is suppletive.[11] Another use of the feature COMPLEXITE = motiv-sem is to identify duplicates such as *thérapique*.A-*thérapeutique*.A 'therapeutic' where the two lexemes have identical meaning and grammatical category.

Finally, combining feature-values allows to account for cross- and back-formations exemplified in the last two rows of Table 6. Cross-formations (Becker, 1993) are simply encoded by means of ORIENTATION = indirect (*bellicisme*.Nm-*bellliciste*.Nm 'warmongering'-'warmonger'). As far as back-formation is concerned (Bauer, 1983; Rainer, 2004; Štekauer, 2015; Manova, 2019) like *hydroplaner*.V-*hydroplanage*.Nm 'to hydroplane'-'hydroplaning' (Namer, 2012), the feature combination is meant to highlight the fact that the verb is derived from a formally more complex noun. This is achieved by the feature set CSTR_1 = X, CSTR_2 = Xage, indicating that the verb is formally simpler than the noun, combined with the feature ORIENTATION = des2as, indicating that *hydroplaner*.V is derived from *hydroplanage*.Nm.

In summary, Démonette's comprehensive set of features and values are used for the description of many types of derivational relations, including those between distant members of the same family. These detailed descriptions enable the identification of families sharing relations with identical properties. These families can be aligned into derivational paradigms (Bonami & Strnadová, 2019). Furthermore, the separation of the feature encoding properties of different levels of representation ensures a description of these paradigms consistent with the ParaDis model proposed by Hathout and Namer (2022).

## 4.3 Table of lexemes

Démonette includes a table of lexemes where the lexeme properties that are independent of the morphological relations are recorded. These properties are categorical, inflectional, phonological, and ontological. Part of the information is recorded in both in the table of lexemes and relations: the written form (i.e., lemma), the grammatical category and the lexeme indices.

---

[11] On the notion of suppletion, see Boyé (2006); on the distinction between base, stem, and theme, see Roché (2010); for a discussion of allomorphy and suppletion, see Paster (2017) and the references cited there.

The table of lexemes contains the vocabulary of all the resources included in Démonette-2. It also includes the entries of the electronic dictionary GLAWI[12] (Sajous et al., 2015), derived from the French Wiktionary. The table ensures the consistency and stability of the database. Stability stems from its lexical coverage, which tends to be complete, so that the addition of new derivational relations or the modification of existing descriptions does not normally affect its content. It contains lexemes in the sense of Matthews (1974/1991), i.e., morphologically simple or complex adjectives, adverbs, nouns, and verbs. It also includes grammatical elements like pronouns, interjections, prepositions, determiners, onomatopoeia, and utterance fragments. They are involved in the analysis of verbs like *vouvoyer*.V 'to address someone using the formal pronoun vous' coined from the pronoun *vous* 'formal 2nd person pronoun'. Similarly, the verb *pschitter*.V 'to spray' is coined from the onomatopoeia *pschit,* the adjective *trentième*.A 'thirtieth' from the determiner (cardinal) *trente.*Num 'thirty', the noun *fortengueulisme*.Nm 'loudmouth behavior' from the fragment *fort en gueule* 'loudmouth', and the nouns *zutisme*.Nm French literary movement at the end of the 19th century that said *zut!* 'damn!' to everything and opposed the very serious Parnassians and *zutiste*.Nm 'member of the Zutisme movement' from the interjection *zut* 'damn'.

Entries in the table of lexemes are identified by a unique index (lexical identifier or LID), and are represented by a written form (i.e., lemma) and a grammatical category.

| LID | Lemme | Cat |
|-----|-------|-----|
| l256593 | mentir 'to lie' | V |
| l256567 | menteur 'lier' | Adj |
| l90317 | abolitioniste 'abolitionist' | Nm |

Table 7. Examples of entries in the table of lexemes. Entries are identified by a lexical identifier (LID), a lemma and a grammatical category

Table of lexemes contains the inflectional paradigms of the verbs, adjectives, and nouns in written form. It also contains their phonemic transcription. The paradigm cells (graphemic and phonemic) are represented by attribute:value pairs where the attribute describes the morphosyntactic features and the value the corresponding word form. The morphosyntactic features are encoded in Multext format (Ide & Véronis, 1994). Verb paradigms have 53 forms (Table 8 only shows 6 of them), adjective paradigms have 4 forms, and noun paradigms have 2 forms.

---

[12]   http://redac.univ-tlse2.fr/lexiques/glawi.html

| Lemma | Cat | Inflected written forms | Phonemic transcription of the inflected forms |
|---|---|---|---|
| mentir 'lie' | V | Vmip1s-:mens; Vmip2s-:mens; Vmmp2s-:mens; Vmii3p-:mentaient; Vmii1s-:mentais; Vmii2s-:mentais; … | Vmip1s-:mã; Vmip2s-:mã; Vmmp2s-:mã; Vmii3p-:mã.tɛ; Vmii1s-:mã.tɛ; Vmii2s-:mã.tɛ; … |
| menteur 'lying' | A | Afpms:menteur; Afpmp:menteurs; Afpfs:menteuse; Afpfp:menteuses | Afpms:mã.tœʁ; Afpmp:mã.tœʁ; Afpfs:mã.tøz; Afpfp:mã.tøz |
| abolitionniste 'abolitionist' | Nm | Ncms:abolitionniste; Ncmp:abolitionnistes | Ncms:a.bɔ.li.sjɔ.nist; Ncmp:a.bɔ.li.sjɔ.nist |

Table 8. Examples of the inflectional paradigms described in the table of lexemes. The paradigms are provided both in written and phonemic form

Verb entries also contain a description of the stem spaces in the form of structured sets of 12 stems following (Bonami & Boyé, 2003; Boyé, 2006; Boyé & Bonami, 2002, 2006). Future versions of Démonette will also include stem spaces of adjectives and nouns. Table 9 shows the stem space of the verb *mentir*.V 'to lie'. The headings are the features of the cells representing the principal parts that each theme allows to reconstruct.

| Ind. Impft | Ind. prs. pl. 3 | Ind. prs. sg | Part. prs | Imp. sg | Imp. pl | Subj. prs. sg | Subj. prs. pl. 12 | inf | Ind. fut | Ind. pst | Part. pst |
|---|---|---|---|---|---|---|---|---|---|---|---|
| mãt | mãt | mã | mãt | mã | mãt | mãt | mãt | mãti | mãti | mãti | mãti |

Table 9. Stem space of the verb *mentir* 'to lie'

The entries for feminine (resp. masculine) nouns that describe animate beings provide the LID of their masculine (resp. feminine) counterparts. These correspondences may be used to complement the table of relations with pairs that contain the correspondent lexemes. Table 10 illustrates these correspondences and the way spelling variants are represented in the table of lexemes.

| Lemma | Cat | LID | gender correspondent | variants |
|---|---|---|---|---|
| abolitioniste | Nf | l90312 | l90313 | l90316 |
| abolitioniste | Nm | l90313 | l90312 | l90317 |
| abolitionniste | Nf | l90316 | l90317 | l90312 |
| abolitionniste | Nm | l90317 | l90316 | l90313 |

Table 10. Description of gender correspondents and of variants in the Table of lexemes. The values in the fields are the LID of the target lexemes.

Moreover, 26% of the common nouns in the table of lexemes are semantically annotated (see the paper by Huguin et al. in this volume). These annotations will be added to the table of lexemes in

the next version of Démonette. The annotation gives the ontological class of these nouns, determined using tests adapted from the FrSemCor project (Barque et al., 2020). The tag set is based on WordNet's Unique Beginners (Miller et al., 1990; Fellbaum, 1998). It includes 22 simple tags (e.g., Entity, Animal, Person, Artifact, Cognition, State, Attribute, Event, Act) and 21 complex ones (e.g., GroupxPerson, Act + Cognition), see Miller (1990).

## 5. The word formations described in Démonette-2

Démonette-2 describes the morphological relations between 111 059 pairs of lexemes. The table of relations therefore contains 222 118 entries since all the relations are symmetric. A subset of derivational processes present in these entries are shown in Table 11. Only processes that form at least 500 complex lexemes in Démonette-2 are listed. The processes are listed as patterns and grammatical categories. Table 11 also gives the number of entries in the table of relations where the patterns occur.

| Pattern | number | Pattern | number | Pattern | number | Pattern | number |
|---------|--------|---------|--------|---------|--------|---------|--------|
| Xion,Nf | 14 290 | Xage,Nm | 14 138 | Xment,Nm | 12 949 | Xeur,Nm | 12 596 |
| Xien,A | 9 655 | Xien,Nm | 9 003 | X,N ↔ X,V | 4 699 | Xeuse,Nf | 5 661 |
| Xique,A | 4 172 | Xisme,Nm | 3 989 | X,Adj ↔ X,V | 2 932 | Xité,Nf | 2 739 |
| Xif,A | 2 619 | Xette,Nf | 2 045 | Xiste,A | 2 032 | Xéen,A | 1 893 |
| Xiste,Nm | 1 843 | Xéen,Nm | 1 781 | Xable,A | 1 742 | Xrice,Nf | 1 639 |
| Xier,Nm | 1 502 | Xerie,Nf | 1 297 | Xier,A | 1 268 | Xaire,A | 1 229 |
| inX,A | 1 097 | Xat,Nm | 1 052 | Xain,A | 1 027 | Xain,Nm | 1 016 |
| Xiste,Nf | 935 | Xiser,V | 879 | déX,V | 771 | X,A ↔ X,N | 767 |
| Xal,A | 710 | reX,V | 624 | Xeux,A | 618 | Xite,Nf | 508 |

Table 11. Sample of WF patterns and categories with the number of entries in the table of relations where they occur. Only patterns with frequency greater than 500 are shown.

We can see in Table 11 that the number of instances of the WF patterns in Démonette-2 correspond globally to their morphological productivity (on the notion of productivity, see the definitions in Bauer (2001, 2005) and the discussion in Dal and Namer (2016); on the productivity of French derivational processes, see Dal (2003); Dal et al. (2008)). In addition to conversion, the most frequent WFs yield suffixed action nouns and agent nouns, noun-based adjectives in *-ique* and *-ien*, property and ideology nouns, diminutives, verbs suffixed in *-iser* and prefixed in *re-, en-* and *dé-*. We can also see that some frequencies are higher than expected because the authors of the resources integrated in Démonette-2 described them extensively. They include the nouns suffixed in *-at* and variants, (Plénat & Roché, 2014) or in *-ier* (Roché, 2004). As a result, the frequency of some WFs does not reflect their productivity. For example, Démonette contains 66 adjectives suffixed in *-ième* (*cent*.Num → *centième*.A, 'hundred'-'hundredth'), 168 feminine nouns suffixed with *-ée* (*arriver*.V → *arrivée*.Nf 'to arrive'-'arrival', *sapin*.Nm → *sapinée*.Nf 'pine tree'-'pine tree plantation') and 94 adjectives prefixed in *anti-* and suffixed in *-ique* (*monarchie*.Nf → *anti-monarchique*.A 'monarchy'-

'anti-monarchy'). This makes Démonette-2 especially interesting since it provides detailed descriptions of the most frequent French WFs (like the suffixation in *-age*) and of less common and more marked processes. On the other hand, some frequencies in Table 11 are lower than expected. For instance, Démonette-2 contains only 879 verbs suffixed in *-iser* and 624 verbs prefixed in *re-*, while both processes are very productive. In comparison, GLAWI contains 4 197 verbs ending in *-iser*. In other words, some morphologically complex verbs are underrepresented because they are not described systematically by any of resources included into Démonette-2. We expect that these biases will diminish as new resources like Glawinette (Hathout et al., 2020; Hathout & Namer, 2021) are added. Glawinette contains 3 459 entries that include a verb ending in *-iser* and 3 785 verbs prefixed in *re-*.

# 6. Online access to the database

Démonette can be accessed and queried online[13]. The site also includes tools intended for specific audiences (e.g., morphologists, speech therapists, teachers) with features designed for their needs (Section 6.4). Note that some features presented in this section are still under development. But overall, the tools already available on the website offer effective and original ways to access the content of the database.

## 6.1 Querying the database

The main function of the online interface is to retrieve lexemes and other units from the table of lexemes and pairs of lexemes from the table of relations. Entries from both tables are selected using queries that combine criteria on forms, part-of-speech, WF patterns, etc. The display of the results is configurable. Users can select the annotations they want to view.

  The database can be queried in several ways. Global searches can be performed on all the fields of a table. For example, searching for "*the entries containing the sequence* `motiv`" in the table of relations retrieves pairs of lexemes where one of their written forms contains the sequence `motiv` (e.g., *motivation*.Nf 'motivation' or *démotiver*.V 'to demotivate'), and the entries where COMPLEXITE is `motiv-sem` or `motiv-form`. Queries can also select entries based on the values of one or more fields. For example, entering `Xisation` in the CSTR_2 field selects 33 entries from the table of relations that meet the condition "*relations in which the second word is suffixed in -isation*" (e.g., *chromer*.V-*chromisation*.Nf 'to chromate'-'chromatation'). In comparison, a global search for "*all relations in which one of the words contains the sequence* `isation`" returns 2 088 pairs, including the 33 previous ones. The result also includes noun-verb and verb-noun pairs where the noun is formed by suffixation in *-ion* on a verb suffixed in *-iser*, (e.g., *canalisation*.Nf-*canaliser*.V 'canalization'-'to canalize') and pairs in indirect complex relations like *salarisation*.Nf-*salarier*.V 'proportion of wage earners in a population'-'to give someone a salaried status').

---

[13]   Démonette's website is https://www.demonette.fr. The code and the update history of the site can be accessed at: https://src.koda.cnrs.fr/llf/web/projects/demonext.

## 6.2 Family graph

Users can access the derivational family of a lexeme in the form of a graph where the relations are tagged by their identifier (RID). The family can be downloaded in tabular format (e.g., CSV) and the graph in graphical format (e.g., PNG). The graphical presentation of the families helps visualize their different subfamilies and apprehend the dynamics of their formation. For example, the family of the noun *paix*.Nf 'peace' (Figure 1) includes the subfamilies of *pacifiste*.N 'pacifist', *appaiser*.V 'to appease', and *implacable*.A 'relentless'.
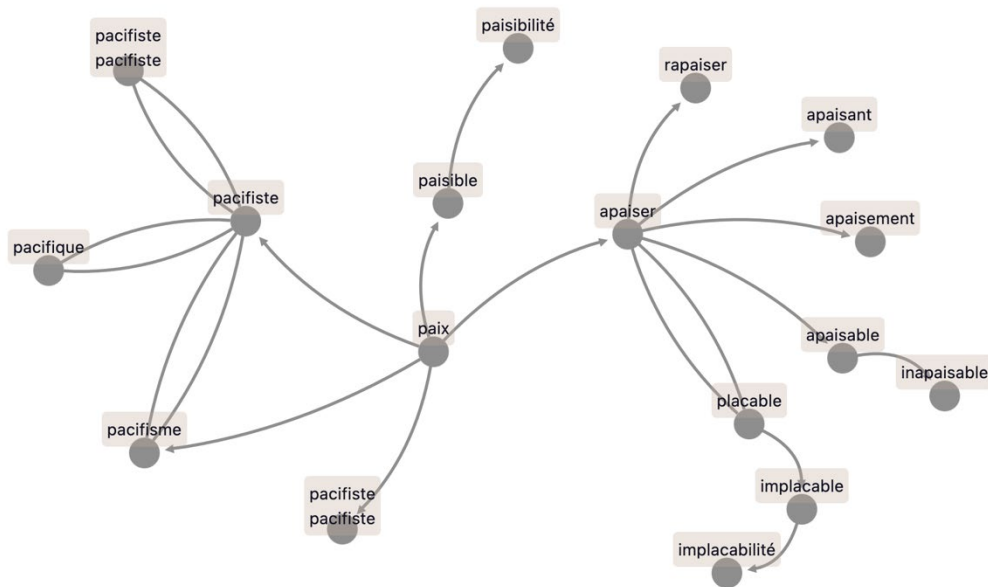


Figure 1. Derivational family of *paix*.Nf 'peace'

## 6.3 Derivational paradigms

An original feature of the Démonette's interface is the possibility to specify graphically the properties of a derivational paradigm and to visualize (and download) the subfamilies that make it up. Consider the example of the paradigm of the triplet *touriste*.Nm-*tourisme*.Nm-*touristique*.Adj 'tourist'-'tourism'-'tourist', shown in Figure 1. A user can retrieve the subfamilies in Démonette that connect a noun suffixed in *-iste*, a noun suffixed in *-isme* and an adjective suffixed in *-ique* using the graphical query tool (navigation menu *specific tools > graph of relations*). The user "draws" the properties of the target structures as a graph pattern like the one on the left hand side in Figure 2, and then uses it as a query. The answers are subfamilies displayed as graphs like the one on the right hand side in Figure 2. This type of queries is designed to look for paradigms. For this reason, only the part of the family that matches the query is displayed.
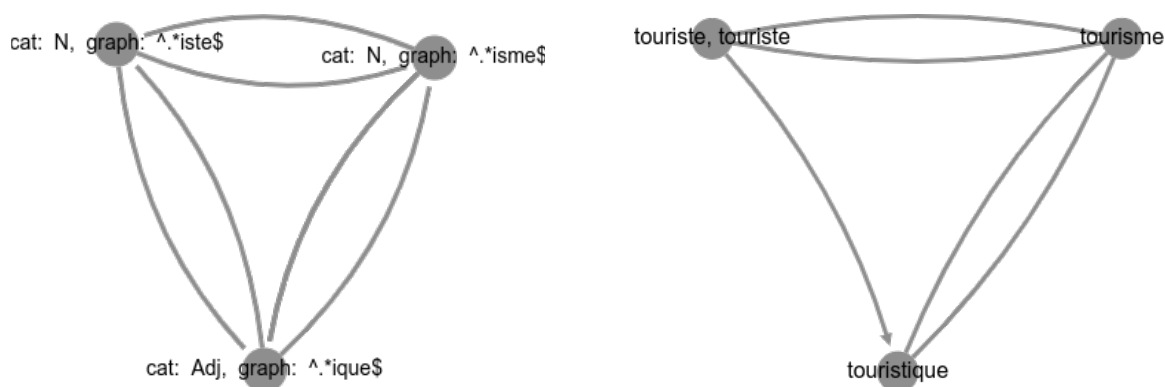
Figure 2. The graph on the left hand side is a query used to retrieve the subfamilies of the derivational paradigm made up of a noun ending in *-isme*, a noun ending in *-iste*, and an adjective ending in *-ique*. The graph on the right hand side is a subfamily that matched the query

As we can see, these queries are much more powerful than the ones presented in Section 6.1 because they are not limited to the properties of pairs of lexemes: graph queries retrieve sets of lexeme pairs. For instance, the query in Figure 2 retrieves 7 triplets of lexemes connected by the same relations. The triplets can be exported in a file in tabular format. Graphical querying is primarily intended for professionals in the field of pedagogy and speech therapy; the representations can possibly be used as illustrative aids for learners and patients, particularly in approaches based on explicit instructions about derived word families.

## 6.4 Specialized tools for speech therapists and for teachers

The website offers nine interactive tools for speech therapy or education (elementary, middle school, and high school) to build resources from which linguistic targets can be selected for the design of assessment tasks and for therapeutic or educational activities. The tools make use of various subsets of the database and possibly other relevant resources. They are complemented by a series of tutorials in the form of video clips[14] that facilitate the use of the online interface, the mastering of the principles of derivational morphology, and of the notions manipulated by the interface. A user's guide to the database tested with speech therapists (Amann, 2023) provides further support. It describes why and how to query the Démonette database in the context of interventions with patients suffering from aphasia. Some tools simplify the querying of the table of relations by non-morphologist users by renaming the features describing the relations and limiting the queries to prefixation or suffixation relations only. Other tools are used for more specific tasks. We present four of them below.

---

[14]    These videos are gathered in the Démonext Youtube channel:
       https://www.youtube.com/channel/UCTaNh1R03KwDE8FzCgSgMhw

### 6.4.1. Pseudo-words

Pseudo-words[15] are often used in psycholinguistic experiments to test lexical and morphological processing mechanisms of learners and patients. Pseudo-words respect the phonotactic and graphotactic rules of the language; they can be pronounced by people (e.g., *fronçaison* or *cheminesque*) just as "real" words. The website gives access to a tool that produces pseudo-words that meet a set of user-defined criteria (e.g., the presence of an affix, the size of the word, the number of syllables). For example, it is possible to generate 3-syllable pseudo-words that end with a suffix like: *causinat, fronçaison, bouborat, commercal, poudroaire, girondée, harappel.*

### 6.4.2 Word families

Users also can retrieve the family of the selected one or more words, either directly or via a set of criteria. For example, they can get the family of *abaissable*.A 'lowerable' which includes *baissage*.Nm 'lowering', *baissement*.Nm 'lowering', *baisser*.V 'to lower', *baisseur*.Nm 'male lowerer', *baissière*.Nf 'swale', *rabais*.Nm 'discount', *rabaissement*.Nm 'lowering', *rabaisser*.Nm 'pull down', *rebaisser*.Nm 'lower again', *abaissable*.A 'lowerable', *abaisse*.Nf 'rolled-out pastry', *abaissement*.Nm 'lowering', *abaisser*.Nm 'lower', *abaisseur*.Nm 'male lowerer'. This feature is useful for teachers in front of their students and for speech therapists in explicit instruction situations with patients (generation or judgment task). It allows them to carefully control derivational relations when preparing word lists.

### 6.4.3 Ancestors and descendants

The interface also includes a tool that provides the words directly and indirectly derived from a base (i.e., its descendants) or the lexemes that are part of the derivation history of a morphologically complex word (i.e., its ancestors or ascendants). Figure 3 shows the descendants of the noun *bouchon*.Nm 'cork'. This feature serves the same purpose as the word family finder: to build controlled lists of words to exercise the morphological skills of students or patients.
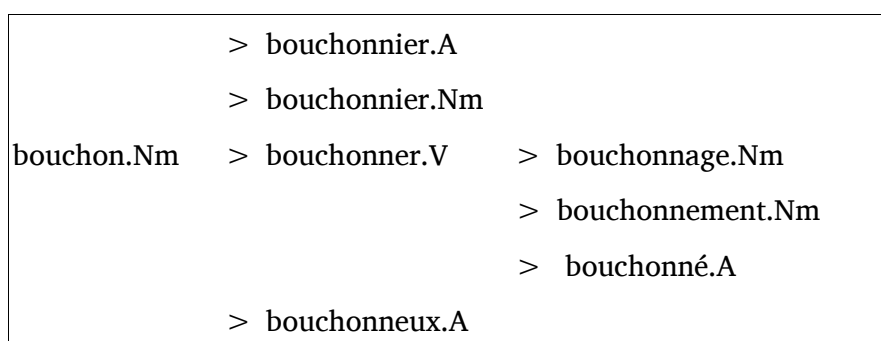
```
                    > bouchonnier.A

                    > bouchonnier.Nm

bouchon.Nm          > bouchonner.V        > bouchonnage.Nm

                                          > bouchonnement.Nm

                                          >  bouchonné.A

                    > bouchonneux.A
```

Figure 3. Descendants of *bouchon*.Nm

---

[15]    A pseudo-word is a string that looks like a real word but is not part of the language's lexicon.

### 6.4.4 Masculine and feminine nouns

Another tool gives access to pairs of masculine and feminine animated nouns, such as *traducteur*.Nm-*traductrice*.Nf 'male translator'-'female translator', or *tigre*.Nm-*tigresse*.Nf, 'tiger'-'tigress'. The tool also provides the masculine and feminine equivalents of animated nouns that are not morphologically related, such as *poule*.Nf-*coq*.Nm, 'hen'-'rooster', and *homme*.Nm-*femme*.Nf 'man'-'woman'. This feature can be used to create exercises for learning vocabulary and noun formation.

## 7. First explorations

The Démonette database has given rise to several works. We briefly describe some of them below.

### 7.1 Anomaly detection based on Formal Concept Analysis

Some of these works aim at the improving Démonette by systematically checking its consistency, automatically identifying errors and gaps, and suggesting solutions to correct them. For example, Juniarta et al. (2022) apply Formal Concept Analysis to the table of relations to identify some of the errors that the database may contain and some of the gaps that could be filled. Démonette-2 was fed using relatively heterogeneous resources (Section 5) and its creation focused on the harmonization of their content; on the other hand, the consistency of the information coming from the different sources has not been checked systematic. As a result, some families may have anomalies, e.g., incorrect or missing relations. These anomalies can be identified by aligning the families to highlight the differences that may exist between them. The method proposed by Juniarta et al. (2022) is based on the description of the families by means of signatures composed of pairs of WF patterns (e.g., `X-Xage`) and lexeme parts-of-speech. The signatures are then placed in a lattice with respect to their inclusion relation. The order is partial because one signature may be included in several others (signatures that contain one or more additional relations). Families that have the same signatures (i.e., whose graphs are homomorphic) can be aligned into paradigms. The alignment can be extended to families whose signatures are partially included in one another.

Presumably, when the signature of a family F is included in the signatures of a set of families such that 95% of them contain an additional relation, then it is likely that this relation is missing in F. This is the case of *enturbannement*.Nm 'enturbanment' which can be added to the family (*turban*.Nm, *enturbanner*.V, *enturbanné*.A) 'turban', 'to enturban', 'enturbaned' on the model of (*mitoufle*.Nf, *emmitoufler*.V, *emmitouflé*.A, *emmitouflement*.Nm) 'glove', 'to muffle', 'muffled', 'mufflement'. Conversely, if the signature of F contains the signatures of a set of families such that 95% of them do not include one of the relations of F, it is likely that this relation is erroneous, such as *palefrenat*.Nm 'stable staff' in the family (*palefroi*.Nm, *palefrenier*.Nm, *palefrenat*.Nm) 'palfrey', 'horse groom', 'stable staff': *palefrenat*.Nm has no equivalent in similar families like (*voiture*.Nf, *voiturier*.Nm) 'car', 'valet'. The anomalies identified in this way may concern the lexemes of a family or the relations between them. When a lexeme is missing, its lemma is predicted from the relations that should connect it to the rest of the family and from the lemmas of the other lexemes in the family. In addition, Juniarta et al. (2022) developed an interface for checking and correcting

families with anomalies. The method was developed and tested on Démonette-1. It will soon be applied to Démonette-2.

## 7.2 Glawinette

Other works aim to extend the coverage of Démonette so that it better reflects the productivity and frequency of the different processes at work in the attested constructed lexicon. As mentioned in Section 5, Démonette has an extensive but uneven coverage because some WFs and phenomena are overrepresented. On the other hand, other WFs are underrepresented. To make Démonette's coverage more even, Hathout et al. (2020) built the Glawinette lexicon using the entries of the GLAWI dictionary (Sajous & Hathout, 2015). The creation of this lexicon is based on the observation that most morphologically complex words are defined by morphological definitions, i.e., by definitions that include another word from their family, as in (1).

(1)    accomplissement = action d'accomplir
       'accomplishment' = 'act of accomplishing'

The method proposed by Hathout et al. (2020) is based on formal analogy (Lepage, 1998; Stroppa & Yvon, 2005; Langlais & Yvon, 2008). Analogy is first used to identify the word pairs that are most likely to be morphologically related (Hathout, 2008, 2011a). For example, *accomplissement*.Nm-*accomplir*.V, 'accomplishment'-'to accomplish' forms an analogy with *assouplissement*.Nm-*assouplir*.V, 'softening'-'to soften'. Conversely, *action*.Nf 'act', the first noun in the definition in (1), and *accomplissement*.Nm do not form a pair that is likely to occur in an analogy with any other pair of lemmas.

In a first step, pairs of related lexemes such as *accomplissement*.Nm-*accomplir*.V are extracted from the definitions. Only pairs that form analogies with at least four other ones are kept. In a second step, the pairs of each analogical series are separated into two series of words. Analogical patterns are then computed for each pair of words of both series. The idea is to characterize the series of words by means of patterns that describe their most characteristic properties, such as `Xissement` for the series of *accomplissement*.Nm. In a third step, the word patterns are aligned to form pairs of patterns. In a fourth step, a linguistically motivated fine-grained alternation pattern is selected for each word pair. The resulting lexicon contains 79 167 lexemes and 161 117 pairs of lexemes. The accuracy of the method is above 99% for the pairs and around 75% for the patterns. The quality of Glawinette allows a relatively easy integration into Démonette (Hathout & Namer, 2021).

## 7.3 Other exploitations

Several other studies based on Démonette are presented in the articles of this volume. For example, Calderone et al. (this volume) present a phonologizer able to predict phonemic transcriptions of lemmas and inflected forms using a neural model trained on the Flexique (Bonami et al., 2014) and GLàFF lexicons. This tool will be used to predict missing transcriptions in Démonette-2: about 10%

of the written forms in the table of lexemes do not have phonemic transcriptions or have transcriptions that do not conform to the IPA standard.

Other papers focus on the use of Démonette in speech therapy. They pursue promising avenues of clinical research that explore intervention models for derivational morphology disorders (as a primary or secondary goal; Galuschka & Schulte-Körne, 2016), the potential of derivational morphology to support important skills such as spelling or vocabulary (as a better predictor of academic success) and the ability to support lexical-semantic mechanisms in children or adults (Goodwin & Ahn, 2010). On the other hand, the state of current knowledge in this area highlights the need to develop methodologically valid tools. The goal is that researchers transfer relevant empirical data to enable clinicians to develop remediation protocols. For example, Duboisdindien, Cattini and Dal (this volume) present a scripted clinical situation in which a speech-language pathologist wishes to develop a derivational morphology intervention aimed at improving the lexical skills of patients with developmental language disorders. Démonette is used in this work to select the relevant targets for the speech therapy intervention. Other work has been done in this direction. We refer to the introduction of the volume for a presentation of other studies that use Démonette.

The possible uses of Démonette are many: in psycholinguistics, speech therapy, NLP, theoretical and descriptive research in the fields of lexicon in general and morphology in particular, vocabulary learning in 4th to 6th grades and morphology teaching at university. It also allows researchers in statistical linguistic to easily create experimental material. A tutorial by Marine Wauquier, Juliette Thuilier and Delphine Tribout, available on the Démonext project website,[16] shows how Démonette can be used in quantitative linguistics. It describes in detail how to study the formation of French demonyms and the competition between the suffixes *(-ais, -éen, -ien, -ois*, etc.) that are used to coin them (Thuilier et al., forthcoming). The tutorial presents the loading of the database, the observation of the tables, the selection of the data according to different criteria such as the presence of labial or alveolar consonants at the end of the toponym from which the demonym derives. It is then possible to observe the distribution of these different properties in relation to the suffixes and to design statistical tests that highlight possible correlations between the properties and the affixes and to estimate their significance. Some trends emerge from these analyses: country names seem to favor the suffix *-ais,* while city names seem to be more often suffixed with *-éen*.

## 8. Conclusion

This paper presents version 2 of the Démonette derivational database created by the members of the Démonext project. Démonette-2 contains a much larger number of entries than the previous versions and describes a much wider range of derivation relations. The way it has been designed and fed allows the base to cover many phenomena that are particularly interesting from a linguistic point of view. They include the suffixation in *-at,* which tend to select learned and suppletive stems,

---

[16] https://www.demonext.xyz/morphologie-et-analyse-statistique/

conversion, whose direction cannot always be determined, and parasynthetic formations, whose formal and semantic motivations are provided by different lexemes in their derivational families. Démonette-2 preserves the distinguishing features of the first versions of the database, namely its relational nature, the separation of the different levels of description (i.e. morphological, formal, categorical, semantic), and the quality of the resources used to feed it. The other important contribution of Démonette-2 is its online interface, designed for a wider public than the users of Démonette-1. Some features of the interface have been designed by speech therapists and psycholinguists to meet the specific needs of these audiences. Both Démonette and its interface[17] have been made publicly available.

Démonette-2 is a long-term project. Future versions will provide access to the original resources. We also plan to complement Démonette's coverage with more generalist resources such as Glawinette, which have a more even coverage of the general lexicon. In the near future, we also plan to integrate the results of the experimental works initiated in the Démonext project.

## References

Amann, E. (2023). *Guide d'utilisation de la base de données morphologiques Démonette2 pour créer des listes contrôlées de mots en orthophonie.* Master d'orthophonie, Université de Lorraine.

Aronoff, M. (1994). *Morphology by Itself.* MIT Press.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM).* Linguistic Data Consortium, University of Pennsylvania.

Batsuren, K., et al. (2022). UniMorph 4.0: Universal Morphology. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference,* Marseille, France (pp. 840–855). European Language Resources Association. https://aclanthology.org/2022.lrec-1.89

Barque, L., Haas, P., Huyghe, R., Tribout, D., Candito, M., Crabbé, B., & Segonne, V. (2020). FrSemCor: Annotating a French corpus with supersenses. In *Proccedings of the Twelfth Edition of its Language Resources and Evaluation Conference – LREC* (pp. 5912-5918). European Language Resource Association (ELRA). https://aclanthology.org/2020.lrec-1.724

Bauer, L. (1983). *English Word-Formation.* Cambridge University Press.

Bauer, L. (1997). Derivational Paradigms. In G. Booij & J. van Marle (Eds), *Yearbook of Morphology 1996* (pp. 243–256). Kluwer.

Bauer, L. (2001). *Morphological Productivity.* Cambridge University Press.

Bauer, L. (2005). Productivity: Theories. In P. Štekauer & R. Lieber (Eds), *Handbook of Word-Formation* (pp.315–334). Springer.

---

[17]   https://www.demonette.fr/

Becker, T. (1993). Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In G. Booij & J. Van Marle (Eds), *Yearbook of Morphology 1993* (pp. 1–27). https://doi.org/10.1007/978-94-017-3712-8_1

Bonami, O., & Boyé, G. (2002). Suppletion and stem dependency in inflectional morphology. In F. van Eynde, L. Hellan & D. Beermann (Eds.), *Proceedings of the 8th International HPSG Conference* (pp. 51–70). CSLI Publications.

Bonami, O., & Boyé, G. (2003). Supplétion et classes flexionnelles. *Langages*, *152*, 103-126. https://doi.org/10.3406/lgge.2003.2441

Bonami, O., & Boyé, G. (2006). Deriving inflectional irregularity. In S. Müller (Ed), *Proceedings of the 13th International Conference on HPSG* (pp. 361–380). CSLI Publications. http://dx.doi.org/10.21248/hpsg.2006.20

Bonami, O., Caron, G., & Plancq, C. (2014). Construction d'un lexique flexionnel phonétisé libre du français. *4e Congrès Mondial de Linguistique Française* (pp. 2583-2596). SHS Web of Conference. https://doi.org/10.1051/shsconf/20140801223

Bonami, O., & Guzmán Naranjo, M. (2023). Distributional evidence for derivational paradigms. In S. Kotowski & I. Plag (Eds), *The Semantics of Derivational Morphology: Theory, Methods, Evidence* (pp. 219–258). De Gruyter.

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

Booij, G. (1986). Form and Meaning in Morphology: the case of Dutch "agent-nouns". *Linguistics*, *24*, 503–518.

Booij, G. (2008). Paradigmatic Morphology. In B. Fradin (Ed), *La raison morphologique. Hommage à la mémoire de Danielle Corbin* (pp. 29–38). John Benjamins Publishers.

Booij, G. (2010). *Construction Morphology*. Oxford University Press.

Booij, G., & Masini, F. (2015). The role of second order schemas in the construction of complex words. In L. Bauer, L. Körtvélyessy & P. Štekauer (Eds), *Semantics of complex words* (pp. 47–66). Springer.

Boyé, G. (2006). Suppletion. In K. Brown (Ed), *Encyclopedia of Language and Linguistics* (2nd Edition, pp. 297-299). Elsevier.

Dal, G. (Ed) (2003). La productivité morphologique en questions et en expérimentations, *Langue Française, 140*.

Dal, G., Fradin, B., Grabar, N., Namer, F., Lignon, S., & Zweigenbaum, P. (2008). Quelques préalables au calcul de la productivité des règles constructionnelles et premiers résultats. *1er Congrès Mondial de Linguistique Française* (pp. 1525-1538) Institut de Linguistique Française. https://doi.org/10.1051/cmlf08184

Dal, G., & Namer, F. (2016). Productivity. In G. Stump, & A. Hippisley (Eds), *The Cambridge Handbook of Morphology* (Chap. 4, pp. 70–89). Cambridge University Press.

Fellbaum, C. (Ed) (1998). *WordNet: An Electronic Database*. MIT Press.

Fradin, B. (2003). *Nouvelles approches en morphologie*. Presses Universitaires de France.

Fradin, B. (2018). Paradigms and the role of series in derivational morphology. *Lingue e linguaggio*, *17*(2), 155–172. DOI: 10.1418/91863

Fradin, B. (2020). Characterizing derivational paradigms. In A. Bagasheva, J. Fernández-Domínguez & C. Lara-Clares (Eds), *Paradigmatic relations in derivational morphology* (pp. 49–84). Koninklijke Brill NV.

Fradin, B. (2021). Caractériser les paradigmes dérivationnels. *Verbum*, *43*, 149-178.

Galuschka, K., & Schulte-Körne, G. (2016). The diagnosis and treatment of reading and/or spelling disorders in children and adolescents. *Deutsches Ärzteblatt International*, *113*(16), 279–286. DOI: 10.3238/arztebl.2016.0279

Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of dyslexia*, *60*(2), 183–208. https://www.jstor.org/stable/23764644

Hathout, N. (2008). Acquisition of the morphological structure of the lexicon based on lexical similarity and formal analogy. *Textgraphs-3*, 1–8.

Hathout, N. (2009). *Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie*. Habilitation à diriger des recherches. Universités de Toulouse II-Le Mirail.

Hathout, N. (2011a). Morphonette: a paradigm-based morphological network. *Lingue e linguaggio*, *2011*(2), 245–264.

Hathout, N. (2011b). Une analyse unifiée de la préfixation en *anti-*. In M. Roché et al. (Eds), *Des Unités Morphologiques au Lexique* (pp. 251–318). Hermès/Lavoisier.

Hathout, N., & Namer, F. (2014a). Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology*, *11*(5), 125–168. https://aclanthology.org/2014.lilt-11.6.pdf

Hathout, N., & Namer, F. (2014b). Discrepancy between form and meaning in Word Formation: the case of over- and under-marking in French. In F. Rainer, W. U. Dressler, F. Gardani & H. C. Luschützky (Eds), *Morphology and meaning* (pp. 177-190). John Benjamins.

Hathout, N., & Namer, F. (2014c). La base lexicale Démonette : entre sémantique constructionnelle et morphologie dérivationnelle. In *Actes de la 21ᵉ conférence annuelle sur le traitement automatique des langues naturelles – TALN-2014* (pp. 208–219). ATALA.

Hathout, N., & Namer, F. (2018). La parasynthèse à travers les modèles : des RCL au ParaDis. In O. Bonami, G. Boyé, G. Dal, H. Giraudo & F. Namer (Eds). *The lexeme in descriptive and theoretical morphology* (pp. 365–399). Language science Press.

Hathout, N., & Namer, F. (2019). Paradigms in word formation: what are we up to? *Morphology, 29*(2), 153–165.

Hathout, N., & Namer, F. (2021). Adding Glawinette into Démonette: practical consequences and theoretical questions. In F. Namer, N. Hathout, S. Lignon, Z. Žabokrtský & M. Ševčíková (Eds). *Third International Workshop on Resources and Tools for Derivational Morphology – DeriMo2021* (pp. 70–75). ATILF.

Hathout, N., & Namer, F. (2022). ParaDis: a Family and Paradigm Model. *Morphology, 32,* 153–195. https://doi.org/10.1007/s11525-021-09390-w

Hathout, N., Sajous, F., & Calderone, B. (2014). GLÀFF, a Large Versatile French Lexicon. *Proceedings of the Ninth International Conference on Language Resources and Evaluation –LREC'1* (pp. 285–298). European Language Resources Association (ELRA).

Hathout, N., Sajous, F., Calderone, B., & Namer, F. (2020). Glawinette: a linguistically motivated derivational description of French acquired from GLAWI. In N. Calzolari, F. Béchet, P. Blache, & K. Choukri, (Eds), *Proceedings of the Twelfth International Conference on Language Resources and Evaluation – LREC 2020* (pp. 3870–3878). European Language Resources Association (ELRA).

Hockett, C. F. (1954). Two Models of grammatical description. *Word, 10,* 210–234.

Iacobini, C. (2020). Parasynthesis in Morphology. In M. Aronoff (Ed), *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.509

Juniarta, N., Bonami, O., Hathout, N., Namer, F. & Toussaint, Y. (2022). Organizing and Improving a Database of French Word Formation Using Formal Concept Analysis. *13th Language Resources and Evaluation Conference – LREC 2022* (pp. 3969–3976). European Language Resource Aassociation (ELRA).

Ide, N., & Véronis, J. (1994). MULTEXT: Multilingual Text Tools and Corpora). In *The 15th International Conference on Computational Linguistics – COLING 1994* (vol. 1, 588-592). https://aclanthology.org/C94-1097.pdf

Koehl, A. (2012). *La construction morphologique des noms désadjectivaux suffixés en français*. Thèse de doctorat en Sciences du Langage, Université de Lorraine.

Koehl, A., & Lignon, S. (2014). Property nouns with *-ité* and *-itude*: formal alternation and morphopragmatics or the sad-itude of the Aité$_N$. *Morphology, 24*(4), 351–376.

Kyjánek, L. (2018). *Morphological Resources of Derivational Word-Formation Relations*. Technical Report, Charles University, ÚFAL. TR-2018-61.

Kyjánek, L., Žabokrtský, Z., Ševčíková, M., & Vidra, J. (2020). Universal Derivations 1.0, A Growing Collection of Harmonised Word-Formation Resources. *The Prague Bulletin of Mathematical Linguistics, 115*, 5–30.

Kyjánek, L., Žabokrtský, Z., Vidra, J., & Ševčíková, M. (2021). *Universal Derivations v1.1*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University. http://hdl.handle.net/11234/1-3247

Jackendoff, R., & Audring, J. (2020). *The Texture of the Lexicon: Relational Morphology and the Parallel Architecture*. Oxford University Press.

Langlais, P., & Yvon, F. (2008). *Scaling up analogy*. Technical Report, Télécom ParisTech.

Lasserre, M., & Montermini, F. (2014). How is the meaning of complex lexemes constructed? A study of neoclassical compounds in *-cratie / -crate* and *-logie / -logue*. *Italian Journal of Linguistics, 26*(2), 157–181.

Lepage, Y. (1998). Solving analogies on words: an algorithm. In *36th Annual Meeting of the Association for Computational Linguistics and 17th international conference on Computational Linguistics* (pp. 728–734). https://aclanthology.org/P98-1120.pdf

Manova, S. (2019). Substractive Morphology. In M. Aronoff (Ed), *Oxford Bibliographies in Linguistics*. Oxford University Press.

Matthews, P. H. (1972). *Inflectional Morphology*. Cambridge University Press.

Matthews, P. H. (1974/1991). *Morphology*. Cambridge University.

McCarthy, A. D., Kirov, C., Grella, M., Nidhi, A., Xia, P., Gorman, K., Vylomova, E., Mielke, S. J., Nicolai, G., Silfverberg, M., Arkhangelskiy, T., Krizhanovsky, N., Krizhanovsky, A., Klyachko, E., Sorokin, A., Mansfield, J., Ernštreits, V., Pinter, Y., Jacobs, C. L., Cotterell, R., Hulden, M. & Yarowsky, D. (2020). UniMorph 3.0: Universal Morphology. *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 3922-3931), European Language Resource Association (ELRA). https://aclanthology.org/2020.lrec-1.483

Miller, G. A. (1990). Nouns in WordNet: A Lexical Inheritance System. *International Journal of Lexicography, 3*(4), 245–264.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Introduction to Wordnet: An On-line Lexical Database. *International Journal of Lexicography, 3*(4), 235–244.

Namer, F. (2009). *Morphologie, Lexique et TAL : l'analyseur DériF*. Hermes Sciences Publishing.

Namer, F. (2012). Nominalisation et composition en français : d'où viennent les verbes composés ? *Lexique, 20*, 173-205.

Namer, F. (2013). A Rule-Based Morphosemantic Analyzer for French for a Fine-Grained Semantic Annotation of Texts. In C. Mahlow & M. Piotrowski (Eds), *International Workshop on Systems and*

*Frameworks for Computational Morphology – SFCM 2013* (pp. 93–115). Springer. DOI: 10.1007/978-3-642-40486-3_6

Namer, F., & Hathout, N. (2020). ParaDis and Démonette – From Theory to Resources for Derivational Paradigms. *The Prague Bulletin of Mathematical Linguistics*, *114*, 5–33.

Namer, F., Lignon, S., & Hathout, N. (2017). Adding morpho-phonological features into a French morpho-semantic resource: the Démonette derivational database. In E. Litta & M. Passarotti (Eds). *Proceedings of the First International Workshop on Resources and Tools for Derivational Morphology – DeriMo* (pp. 49–61). EDUCatt.

Namer, F., & Villoing, F. (2014). Sens morphologiquement construit et procédés concurrents : les noms de spécialistes en *-logue* et *-logiste*. *Revue de Sémantique et Pragmatique*, *35-36*, 7-26.

New, B. (2006). Lexique 3 : Une nouvelle base de données lexicales. *Actes de la Conférence Traitement Automatique des Langues Naturelles – TALN 2006*. Louvain, Belgique.

Paster, M. (2017). Alternations: Stems and Allomorphy. In A. Hippisley & G. Stump (Eds), *The Cambridge Handbook of Morphology* (pp. 93–116). Cambridge University Press.

Plénat, M., & Roché, M. (2014). La suffixation dénominale en *-at* et la loi des (sous-)séries. In F. Villoing, S. David & S. Leroy (Eds). *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux* (pp. 47-74). Presses Universitaires de Paris Ouest.

Rainer, F. (2004). La retroformazione come fenomeno analogico. In M. Grossmann & F. Rainer (Eds). *La formazione delle parole in italiano* (pp. 495–497). Niemeyer.

Robins, R. H. (1959). In defence of WP. *Transactions of the Philological Society*, *58*(1), 116–144.

Roché, M. (2004). Mot construit ? Mot non construit ? Quelques réflexions à partir des dérivés en *-ier(e)*. *Verbum*, *26*(4), 459-480.

Roché, M. (2010). Base, thème, radical. *Recherches Linguistique de Vincennes*, *39*, 95-133.

Roché, M. (2011a). Pression lexicale et contraintes phonologiques dans la dérivation en *-aie* du français. *Linguistica*, *51*, 5-22.

Roché, M. (2011b). Quel traitement unifié pour les dérivations en *-isme* et en *-iste*. In M. Roché et al. (Eds), *Des Unités Morphologiques au Lexique* (pp. 69-143). Hermès/Lavoisier.

Roché, M. (2017a). *Les familles dérivationnelles : comment ça marche ?* (ms.)

Roché, M. (2017b). *Un exemple de réseau constructionnel : ethniques, toponymes, gentilés.* Université Toulouse-Jean Jaurès.

Roché, M., & Plénat, M. (2012). Tous les déverbaux en *-at* sont-ils des conversions du thème 13 ? *3ᵉ Congrès Mondial de Linguistique Française – CMLF* (pp. 1387-1405). SHS Web of Conference. https://doi.org/10.1051/shsconf/20120100121

Romary, L., Salmon-Alt, S., & Francopoulo, G. (2004). Standards going concrete: from LMF to Morphalou. In *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries* (pp. 22-28). *COLING*. https://aclanthology.org/W04-2104.pdf

Sajous, F., Hathout, N., & Calderone, B. (2013). GLàFF, un gros lexique à tout faire du français. *Actes de la 20ᵉ conférence sur le Traitement Automatique des Langues Naturelles – TALN'2013* (pp. 285-298). http://talnarchives.atala.org/TALN/TALN-2013/taln-2013-long-021.pdf

Sajous, F., & Hathout, N. (2015). GLAWI, a free XML-encoded Machine-Readable Dictionary built from the French Wiktionary. *Proceedings of the eLex 2015 conference* (pp. 405–426). https://shs.hal.science/halshs-01191012/

Štekauer, P. (2014). Derivational Paradigms. In R. Lieber & P. Štekauer (Eds), *The Oxford Handbook of Derivational Morphology*. Oxford University Press, 354–369.

Štekauer, P. (2015). Backformation. Word-Formation. In P. O. Müller, I. Ohnheiser, S. Olsen, & P. Štekauer (Eds), *An International Handbook of the Languages of Europe* (vol. 1, pp. 340–352) de Gruyter.

Strnadová, J. (2014). *Les réseaux adjectivaux. Sur la grammaire des adjectifs dénominaux en français.* Thèse de doctorat. Université Paris Diderot & Univerzita Karlova.

Stroppa, N., & Yvon, F., (2005). An Analogical Learner for Morphological Analysis. *Proceedings of the Ninth Conference on Computational Natural Language Learning – CoNLL-2005* (pp. 120–127) Ann Arbor.

Stump, G. (2017). The Nature and Dimensions of Complexity in Morphology. *Annual Review of Linguistics, 3*(1), 65–83.

Thuilier, J., Tribout, D., & Wauquier, M. (forthcoming). Affixal rivalry in French demonyms formation: The role of linguistic and non-linguistic parameters. *Word Structure*.

Tribout, D. (2010). *Les conversions de nom à verbe et de verbe à nom en français.* Thèse de doctorat, Université Paris 7.

Tribout, D. (2012). Verbal stem space and verb to noun conversion in French. *Word Structure, 5*(1), 109–128.

Tribout, D. (2020). Nominalization, verbalization or both? Insights from the directionality of noun-verb conversion in French. *Zeitschrift für Wortbildung / Journal of Word Formation, 4*(2), 187–207.

Vidra, J., Žabokrtský, Z., Ševčíková, M., & Kyjánek, K. (2019). DeriNet 2.0: Towards an All-in-One Word-Formation Resource. In M. Ševčíková, Z. Žabokrtský, E. Litta & M. Passarotti (Eds), *Proceedings of the Second International Workshop on Resources and Tools for Derivational Morphology – Derimo 2019* (pp. 81–90). https://aclanthology.org/W19-8510.pdf

Žabokrtský, Z., Bafna, N., Bodnár, J., Kyjánek, L., Svoboda, E., Ševčíková, M. & Vidra, J. (2022). Towards Universal Segmentations: UniSegments 1.0. *Proceedings of the 13ᵗʰ Conference on Language*

*Resources and Evaluation – LREC 2022* (pp. 1137-1149). https://aclanthology.org/2022.lrec-1.122.pdf