Adding Morpho-phonological Features into a French Morpho-semantic Resource: the Demonette Derivational Database

Fiammetta Namer[(a)], Nabil Hathout[(b)] & Stéphanie Lignon[(a)]

[(a)]UMR 7118 ATILF, CNRS & Univ de Lorraine, France

[(b)]UMR 5263 CLLE-ERSS, CNRS & Univ de Toulouse-Jean Jaurès, France

E-mail: {fiammetta.namer, stephanie.lignon}@univ-lorraine.fr

nabil.hathout@univ-tlse2.fr

Abstract

Demonette (Hathout & Namer [13]) is a derivational database (DDB) of French with a relational structure: its entries describe a large number of properties of derivational relations connecting word pairs, such as LANCER 'launch'→ LANCEUR 'launcher' or LANCEUR → LANCEMENT 'launching'. The entries also specify the categorical, semantic and morpho-phonological properties of the connected words . We here present the morpho-phonological ones and show how Demonette's organization allows an original representation of these properties. Demonette's entries provide phonological transcriptions of the word pairs and syllabic decompositions. It also specifies their stems and the possible variations they display.

1 Introduction

Demonette (Hathout & Namer [13]) is a derivational database (DDB) of French which represents the morphological information in an original way: entries do not describe the properties of the derivatives; they describes the properties of the derivational relations connecting pairs of lexemes, such as LANCER 'launch'→ LANCEUR 'launcher$_{masc}$' or LANCEUR → LANCEMENT 'launching'. These relations specify the derivational properties of the lexemes they connect. One consequence of this conception is that the overall properties of a lexeme are the outcome of all the properties induced by each of the relations the lexeme occurs in. More generally, Demonette's structure is completely determined by this conception: The DDB is redundant, because relations are direct, indirect and bi-directional. Demonette describes relations between derivationally related pairs of lexemes [L1, L2], where L1 is morphosemantically motivated by L2. It includes relations between derived words and their bases (e.g. [LANCEUR, LANCER], where LANCEUR's meaning can be defined as "the one who performs the action of LANCER"), and

relations between base words and their derivatives (eg. [LANCER, LANCEUR], where LANCER means "doing what a LANCEUR does"). The network also contains *indirect* relations between lexemes of the same derivational family, where none is the base of the other such as [LANCEMENT, LANCEUR]. In this relation, LANCEMENT can be defined as "activity performed by the LANCEUR". The relation is part of a network which contains [LANCER, LANCEUR], [LANCEUR, LANCER], [LANCEMENT, LANCER] (LANCEMENT is the "activity of LANCER"), [LANCER, LANCEMENT].

Derivational relations define derivational families, and are organized into paradigms. In previous publications, we focused on the morphosemantic characteristics of Demonette. We here address the morphophonological aspects of the DDB, and we show how these properties are described in Demonette and how morphophonological paradigms can be represented.

2 Derivational databases

One key feature of derivational morphology is its lexicality. Moreover, the analysis of complex lexemes relies on a large amount of memorized information.

In recent years, several efforts have improved the morphological analysis by using large corpora (Cotterrel [9], Lazaridou [15]), but progress on morphological information storage and harmonization has been weaker. A lot remains to be done: the accumulation of morphological knowledge is crucial for many researches in descriptive morphology, lexicology, teaching, etc.

The first DDBs where designed by psycholinguists in order to create experimental data. The best-known DBB is CELEX (Baayen *et al.* [2]) whose first version was released in the 90s. This resource covers English, German and Dutch and offers a broad range of phonological, morphosyntactic, inflectional and derivational information. It remains a reference with no real equivalent, despite its limited coverage, when compared to the size of the corpora available today.

Other large-scale resources have been created for English, such as CatVar (Habash & Dorr [11]), a lexicon of derivational family intended primarily to NLP applications. More recently, a similar resource has been developed for German: DerivBase (Zeller *et al.* [30]) was automatically built from corpora, with the help of distributed semantics methods. Another significant resource is DerivaTario (Talamo *et al.* [26]), a derivational dictionary of Italian; It provides analyses based on strong hypotheses regarding allomorphy and suppletion. For instance, BELLICOSO 'bellicose' is analyzed as a derivative of GUERRA 'war'. For French, the only comparable resource is Demonette. Its main characteristics are presented hereafter.

3 Demonette

One goal of Demonette (Hathout & Namer [13]) is to help satisfy the need for reliable and broad-coverage morphological resources of French. Demonette is a DDB characterized by an original structure based on the derivational relations. Moreover, it can host morphological descriptions from research works such as PhDs in morphology, or from manual-assessed NLP lexical resources, like VerbAction (Tanguy & Hathout [27]). In its current state (Hathout & Namer [14]), Demonette (version 1.3) includes information coming from four sources: DériF (Namer [18], [19]), Morphonette (Hathout [12]), VerbAction and Lexeur (Fabre *et al*. [10]). They have been added in three successive stages. Overall, Demonette contains 167,369 entries. Derived words in Demonette can be deverbal action nouns (ESSORAGE 'spin'), deverbal masculine or feminine agent nouns (RAMASSEUR 'collector', RAMASSEUSE 'collector') or deverbal adjectives (PRODUCTIF 'productive'). Demonette also includes simplex verb predicates (CONSTRUIRE 'build').

The fields used to describe the derivational relations in the Demonette do not form a closed list and can be extended if needed. Among the existing fields, the most original ones are probably those used for the semantic description, and include morphosemantic types (eg. @AGF for feminine agents), concrete definitions giving the meaning of L1 with respect to L2 (eg. MARCHEUSE in the relation [MARCHEUSE, MARCHER] is defined as "she who performs the action of MARCHER"), and abstract definitions generalizing the concrete ones where the meanings of L1 and L2 are replaced by their respective semantic type (eg. @AGF: "she who performs @"). Relations with the same abstract definition are inserted into the same morphosemantic paradigm. This is the case with the ones listed in Table 1.

L1, cat	**L2, cat**	**Type L1**	**Type L2**	**Concrete def**	**Abstract def**	**Affix L1**
MARCHEUSE, N_{Fem} 'walker$_{(fem)}$'	MARCHER, V 'walk'	@AGF	@	"she who performs the action of marcher"	"she who performs the action of @"	euse
ENSEIGNANTE, N_{Fem} 'teacher$_{(fem)}$'	ENSEIGNER ,V 'teach'	@AGF	@	"she who performs the action of enseigner"		ante
DIRECTRICE, N_{Fem} 'director$_{(fem)}$'	DIRIGER,V 'direct'	@AGF	@	"she who performs the action of diriger"		rice

Table 1: Concrete and abstract definitions of three feminine agent nouns

4 Morpho-phonological descriptions within Demonette

The 167,369 [L1, L2] entries of Demonette1.3 have been completed with morphophonological information: L1 and L2 phonological representations, the properties of their stems and exponents, and a description of the morphophonological variations that occur in the [L1, L2] relation. This information is mostly unpredictable in a language with rich morphology such as French and is therefore crucial for a comprehensive description of its derivational system. This additional knowledge is interconnected with the rest of the entries properties and in particular with the morphological and the morphosemantic ones.

In Demonette, morphophonological properties are described in a similar way to the morphosemantic ones: we distinguish concrete and abstract levels; some of the morphophonological descriptions of the L1 and L2 lexemes are induced by the derivational relation which connects them. Morphophonology is both easier and harder to describe than morphosemantics. On the one hand, it is simpler, because IPA transcriptions are part of the mainstream in linguistics: we don't have a similar standard for morphosemantic representation. On the other hand, it is more complex, because lexemes are abstract objects that do not have formal properties by themselves, unlike the inflected forms (or word forms) that realize them. We also consider that each word form can be decomposed into an inflectional stem and an inflectional exponent (Baerman *et al*. [3]).

Following Boyé [8], Bonami & Boyé [4] and Montermini & Bonami [17], we define word form exponents in French as the maximal rightmost strings that are common across the patterns, and interpret all the remaining variation as stem allomorphy (see Spencer [25] and Bonami & Boyé [7] for a discussion), where stems are pure forms (or morphomes, in Aronoff's terms [1]). As often discussed in the literature (see Bonami & Boyé [4, 6], Montermini & Boyé [16], Montermini & Bonami [17] among others), stems form a paradigmatic organization called *stem space* (deriving from Pirrelli & Battista's [21] 'Overall Distribution Schema'). Stem spaces are made of cells forming a graph where stems are in a dependency relation with each other. The value of a stem occupying one cell depends on the value of stem in one or several other cells. By default, this value is inherited from them without change. Allomorphic stems correspond to override of the default inheritance. The complexity of the stem space is language and part-of-speech dependent. For instance, the stem space of French verbs is a graph of at least 13 cells. Table 2 lists the 13 stems of the verb BOIRE 'drink'. Each stem is used by one or several inflection rules to produce one or several forms of the verb[1].

[1] C1 is used for the IND.PRS.SG; C2: IND.PRS.3PL; C3: IND.IPFV & IND.PRS.1PL & 2PL; C4: PTCP.PRS; C5: IMP.2SG; C6: IMP.1PL & 2PL; C7: SBJV.PRS.SG & 3PL; C8: SBJV.PRS.1PL & 2PL; C9: INF.PRS; C10: IND.FUT & COND.PRS; C11: IND.PST; C12: PTCP.PST.M; C13: PTCP.PST.F.

C1	C2	C3	C4	C5	C6	C7	C8	C9
bwa	bwav	byv	byv	bwa	byv	bwav	byv	bwar
C10	**C11**	**C12**	**C13**					
bwa	by	by	by					

Table 2: Stem space of the verb BOIRE 'drink'

Likewise, adjective and noun stems are organized in stem spaces: in French, a 3 cells space is required for adjectives (C1: M.SG; C2: F.SG; C3:M.PL; see Boye & Bonami [5]) and a 2 cells one for nouns (C1:SG; C2:PL; see Roché [23]). Table 3 shows the spaces of the adjective BEAU 'beautiful' and of the noun CHEVAL 'horse'.

ADJ		
C1	**C2**	**C3**
bo	bɛl	bo

NOUN	
C1	**C2**
ʃəval	ʃəvo

Table 3: Stem spaces of the French adjective BEAU and noun CHEVAL.

Stem spaces also play a central role in word formation: word formation patterns use particular cells in the stem space of the input lexemes. For instance, *-able* suffixed deverbal adjectives are formed with the C3 verb stem. Therefore, the stem /byv/ is selected to derive BUVABLE 'drinkable' from BOIRE. Similarly, deadjectival prefixed verbs are generally built on the C2 adjective stem: EMBELLIR, for instance selects the /bɛl/ stem of the adjective BEAU.

In Demonette we only provide the morphophonological properties of lexemes (or more precisely, of the wordforms that realize them) relevant for word formation (Plénat [22], Roché [23]). Therefore, stems and exponents are listed in the [L1, L2] description only if they are involved in derivational constructions. For French, this means that, out of the stem spaces illustrated in Tables 2 and 3, only the following are needed:

- For nouns, C1, e.g. CHEVAL 'horse' → CHEVALIER /ʃəvalje/ 'horseman'
- For adjectives, C1 and C2 are relevant: the M.SG stem /bo/ of BEAU is used to form the property noun BEAUTÉ /bote/ 'beauty', and the F.SG stem /bɛl/ to form the pejorative noun BELLÂTRE /bɛlɑtʀ/ 'fop'.
- Six stems are required for verbs: C1, C4, C12 and C13 are used by V-to-N conversion patterns (C1: SOUTENIR 'support$_V$' → SOUTIEN /sutjɛ̃/ 'support$_N$', C4: COURIR 'run' → COURANT /kurɑ̃/ 'flow', C12: DEVOIR 'owe' → DÛ /dy/ 'due', C13: MÉPRENDRE 'be mistaken' → MÉPRISE /mepriz/ 'mistake', cf. Tribout [28]), C2, used in *-ment* suffixed deverbal event nouns (SOULEVER 'lift$_V$' → SOULÈVEMENT

/sulɛvmã/ ‘lift_N’), and C3 for *-able* suffixed adjectives (BOIRE ‘drink’ → BUVABLE /byvabl/ ‘drinkable’).

The main source of automatic acquisition for the IPA transcription of the selected stems is the freely available database GLÀFF (Sajous *et al*. [24]), which contains more than 1.4 million entries of inflected forms annotated with phonetic representation encoded in SAMPA (Wells [29]). When needed, it is completed with data coming from Lexique3 (New [20]), which uses phonetic transcriptions very similar to SAMPA, which makes the mapping task relatively trivial.

All but one of the stems of the lexemes present in GLÀFF or Lexique3 can be directly retrieved from the word forms for which they have been used. The exception is C3 for verbs, because this stem is always concatenated to an exponent in the word forms: the C3 stem is thus computed from the IND.PRS.1PL form by stripping off the final /õ/ exponent (eg. *buvons* ‘(we) drink’, /byvõ/ = /byv/ ⊕ /õ/). The entries also contain various other pieces of information that describe the morphophonological specificity of L1, L2 and the [L1, L2] relation (see Tables 4 and 5).

In Table 4, the features *Rad1* and *Rad2* can be compared to determine the formal distance between L1 and L2. When [L1, L2] are in a base/derivative relation, as in [BOIRE, BUVEUR], *Rad2* is obtained by removing the suffix *Suf2* (e.g. /œr/) from the word form of the derivative (e.g. BUVEUR). *Rad1* is selected from the stem space of the base (e.g. BOIRE) in such a way that it is the most similar to one of the possible values of *Rad2*. In the example [BOIRE, BUVEUR], it is C3 (see Table 2).

When L1 and L2 are in an indirect relation, as in [ADMIRATEUR_N, ADMIRATION_N], both words being derived from ADMIRER_V ‘admire_V’, the value of Rad_i is obtained by depriving L_i from the suffix Suf_i. For [ADMIRATEUR_N, ADMIRATION_N], we get *Rad1* = /admirat/ and *Rad2* = /admiras/. For each L_i, the Rad_i description also includes the number of syllables $Size_i$, and the properties of its last syllable (onset, vowel, final consonant), as value of, respectively, $LastOnset_i$, $LastV_i$ and $LastC_i$.

L1						**L2**					
Rad1	**Size1**	**LastOns1**	**LastV1**	**LastC1**	**Suf1**	**Rad2**	**Size2**	**LastOns2**	**LastV2**	**LastC2**	**Suf2**
BOIRE ‘drink_V’						BUVEUR ‘drinker (masc)’					
byv	1	b	y	v	--	byv	1	b	y	v	œr
ADMIRATEUR ‘admirer (masc)’						ADMIRATION ‘admiration’					
admirat	3	r	a	t	œr	admiras	3	r	a	s	jõ

Table 4: Rad1 and Rad2

Table 5 gives more examples of [L1, L2] formal properties, and shows how allomorphy is described.

		L_i	Rad_i	Suf_i	Z	Alternation	Concrete Phon Rel	Abstract Phon Rel	Phon Rule
1	L1 L2	boire buveur	byv byv	-- œr	byv	=	byv / byvœr	Z / Zœr	=
2	L1 L2	admirateur admiratif	admirat admirat	œr if	admirat	=	admiratœr / admiratif	Zœr / Zif	=
3	L1 L2	admirer admirateur	admir admirat	-- œr	admir	ǀ at	admir / admiratœr	Z / Zatœr	NONE
4	L1 L2	admirateur admiration	admirat admiras	œr jõ	admira	t ǀ s	admiratœr / admirasjõ	Ztœr / Zsjõ	→ [+sib]
5	L1 L2	extincteur extinction	ekstɛ̃kt ekstɛ̃ks	œr jõ	ekstɛ̃k	t ǀ s	ekstɛ̃ktœr / ekstɛ̃ksjõ	Ztœr / Zsjõ	→ [+sib]
6	L1 L2	éteindre extincteur	etɛ̃ ekstɛ̃kt	-- œr	e	--	etɛ̃ / ekstɛ̃ktœr	etɛ̃ / ekstɛ̃ktœr	NONE
7	L1 L2	aliment 'food' alimentaire 'alimentary'	alimã alimãt	-- ɛr	alimã	ǀ t	alimã / alimãtɛr	Z / Ztɛr	+C
8	L1 L2	cheval 'horse' hippique 'equine'	ʃəval ip	-- ik	--	--	ʃəval / ipik	ʃəval / ipik	NONE

Table5: Identity and variation in a derivational relation

Columns L_i, Rad_i, Suf_i contain the orthographic representation, the radical and suffix of each of the lexemes L1 and L2 (cf. above Table 4).
The other columns describe the properties of the [L1, L2] relation.

- The field ***Concrete Phon(ological) Rel(ation)*** reproduces the sequences formed by the concatenation of *Rad1* and *Rad2* (cf. Table 4).
- The longest common subsequence of *Rad1* and *Rad2* is given in ***Z***. *Z* can be identical to *Rad1* and *Rad2* as with /byv/ and /admirat/, in raw 1 and 2; it can be identical to *Rad1* and included in *Rad2*: both *Rad1* /admir/ in raw 3 and /alimã/ in raw 7 are included in their corresponding *Rad2*; it can be a subpart of *Rad1* and *Rad2*, such as /admira/ (raw 4), /ekstɛ̃k/ (raw 5) and /e/ (raw 6), or it be empty (raw 8).

- When *Z* has a non-null value, the difference between *Rad1* and *Rad2* is given in the field ***Alternation***. When *Rad1* and *Rad2* are identical, the value of *Alternation* is '=' (raws 1 and 2). The value 'tls' (in raws 4 and 5) says that the variation between *Rad1* and *Rad2* is a change in their last consonant; the value 'lat' in raw 3 (resp. 'lt' in raw 7) says that *Rad2* is the concatenation of *Rad1* with /at/ (resp. with /t/). We consider the *Alternation* value to be not relevant (value '--') when the two stems are completely different (raw 8), or when their difference (*i*) is not reproduced elsewhere in the lexicon, and (*ii*) is greater than their likeliness (raw 6).
- When relevant, *Alternation* is characterized phonologically. The explanation (assibilation, insertion, sonorization, etc.) is encoded as a rule in the ***Phonological Rule*** field (last column). The rule is identity, symbolized by '=' in raws 1 and 2. It contains the value 'NONE', e.g. in raw 3 because the difference between *Rad1* /admir/ and *Rad2* /admirat/ does not have a phonological origin but an historical one (/admirat/ is the Latinate bound stem of the verb ADMIRER). Likewise, the rules in raws 6 and 8 have a 'NONE' value because the stem variations between L1 and L2 are not phonologically motivated. Conversely, the 'tls' alternation in the relations of raws 4 and 5 can be qualified as a case of palatalization (or sibilantization), represented by the '→[+sib]' rule. The insertion of /t/ at the stem/suffix boundary of /alimã**t**ɛr/ in raw 7 is phonologically motivated (as opposed to the /at/ insertion in raw 3): it is the sonorization (symbolized with '+C') of the latent final consonant on the orthographical form *aliment*.
- *Z* is used to generalize the *Concrete Phon Rel* into an ***Abstract Phon(ology) Rel(ation)*****,** where the Z symbol substitutes for the value of the *Z* attribute. This abstract relation emphasizes the morphophonological organization of the lexicon, in particular in terms of stem and exponent variation. This abstract representation also identifies the set of morphophonological relations that connect each lexeme to the rest of its derivational family.

The descriptions exemplified in Tables 4 and 5 allows us to separate the derivational relations into four categories according to morphophonological criteria, based on their identity, the variation between their stems, and the nature of their formal relation. This categorization uses the values of *Alternation* and *Abstract Phon Rel* fields. The four categories are:

(i) no stem variation (raw 1, 2);
(ii) phonologically motivated variation (raws 4, 5, 7);
(iii) stem variation surfacing as an alternation not phonologically motivated (raw 3);

(iv) suppletion, i.e. no sequence in common (raws 6, 8).

The alternations define morphophonological classes of derivational (sub-)families: for instance, the same set of A ↔ B stem variations are shared by [L1, L2] pairs in several derivational families, as shown in Table 6. Stem variations are evidenced by the *Abstract Phon Rel* value in each of the relevant [L1, L2] entries.

Table 6 shows that (COMPOSER$_V$ 'compose', COMPOSITEUR$_N$ 'composer (m)', COMPOSITRICE$_N$ 'composer (f)', COMPOSITION$_N$ 'composition') and (INHIBER$_V$ 'inhibit', INHIBITEUR$_N$ 'inhibitor (m)', INHIBITRICE$_N$ 'inhibitor (f)', INHIBITION$_N$ 'inhibition') share the same set of stem variations, and have the same suffix exponents *-eur, -rice* and *-ion*. Moreover, the indirect relations in Demonette highlights the formal organization of the lexicon. These relations make it possible to identify sub-regularities, for instance between EXTINCTEUR$_N$ 'extinguisher' and EXTINCTION$_N$ 'extinction' (raw 5, Table 5) or between PRÉDATEUR$_N$ 'predator' and PRÉDATION$_N$ 'predation': whereas the standard derivational connections between the first noun pair can be retrieved from their individual relations with their verb base ÉTEINDRE 'extinguish', as shown in Table 5, raw 6, there is no such direct base/derived relation in the French contemporary lexicon, between PRÉDATEUR or PREDATION and a common verb base.

		PRED(V)	M. AGENT(N)	F. AGENT(N)	EVENT(N)
Deriv. families		COMPOSER	COMPOSITEUR	COMPOSITRICE	COMPOSITION
		…			
		INHIBER	INHIBITEUR	INHIBITRICE	INHIBITION
A ↔ B	Z ↔ Zit	A	B		
		A		B	
	Z ↔ Z		A	B	
	Z ↔ Zis	A			B
	Zt ↔ Zs		A		B
				A	B

Table 6: Morphophonological organization of derivational families

5 Paradigmatic view of the derivational lexicon

With the organization we outlined above, Demonette has a triple network of morphological, morphosemantic and morphophonological relations able to capture paradigmatic regularities and sub-regularities at different levels. Just like morphosemantics, morphophonological information is described at two levels, a concrete one and an abstract one, which multiplies the perspectives of observation.

For instance, at the concrete level, noun pairs EXTINCTEUR ↔ EXTINCTION, ADMIRATEUR ↔ ADMIRATION and PRÉDATEUR ↔ PRÉDATION

behave in the same way, whereas at the abstract level, (ADMIRER, ADMIRATEUR, ADMIRATION) and (ÉTEINDRE, EXTINCTEUR, EXTINCTION) belong to two distinct series.

Examined at different levels, the same data leads to different findings. For example, crossing morphology and morphophonology leads to the insertion of [PRÉDATEUR, PRÉDATION] in the sub-paradigm of the paradigm (ADMIRATEUR, ADMIRATION, ADMIRER).

If we consider the morphosemantic / morphophonology opposition, triplets (ADMIRER, ADMIRATEUR, ADMIRATION) and (CONSPIRER 'conspire$_V$', CONSPIRATEUR 'conspirator(m)$_N$', CONSPIRATION 'conspiracy$_N$') belong to two different morphosemantic paradigms (ADMIRER and ADMIRATION are stative predicates, whereas CONSPIRER and CONSPIRATION are eventive ones), but to the same morphophonological paradigm; conversely (ENSEIGNER 'teach', ENSEIGNANT 'teacher(m)', ENSEIGNEMENT 'teaching') is in the same morphosemantic paradigm as (CONSPIRER, CONSPIRATEUR, CONSPIRATION), but the two sub-families belong to distinct morphophonological paradigms.

Finally, the two families presented in Table 6 illustrate a case of uniform paradigm: members of the same morphophonological category share the same semantic category and the same part-of-speech (INHIBER and COMPOSER are verbal predicates, COMPOSITEUR and INHIBITEUR, masculine agent nouns, INHIBITRICE and COMPOSITRICE, feminine agent nouns, and COMPOSITION and INHIBITION event nouns). They result from the same derivational processes (the verbs are simplex, and the nouns are suffixed in *-eur*, *-rice* and *-ion* respectively) and are two by two in the same phonological relations, as shown in Table 6.

References

[1] Aronoff, Marc. *Morphology by Itself*, MIT Press: Cambridge, 1994.

[2] Baayen, R. Harald, Piepenbrock, Richard, and Gulikers, Leon. *The CELEX lexical database* (release 2). Linguistic Data Consortium, Philadelphia, 1995.

[3] Baerman, Matthew, Corbett, Greville, Brown, Dunstan and Hippisley, Andrew. *Deponency and Morphological Mismatches*, British Academy and Oxford University: Oxford, 2007.

[4] Bonami, Olivier and Boyé, Gilles. Supplétion et classes flexionnelles dans la conjugaison du français, pp. 102–126. *Langages* 152, 2003.

[5] Bonami, Olivier and Boyé, Gilles. Construire le paradigme d'un adjective, pp. 77–98. *Recherches Linguistiques de Vincennes* 34, 2005.

[6] Bonami Olivier and Boyé Gilles. Remarques sur les bases de la conjugaison. In Delais-Roussarie, Elisabeth and Labrune, Laurence (Eds.), *Des sons et des sens. Données et modèles en phonologie et en morphologie*, pp. 77–90. Hermès-Lavoisier: Paris, 2007.

[7] Bonami, Olivier and Boyé Gilles. De formes en thèmes. in Villoing, Florence, Leroy, Sarah and David, Sophie (eds.), *Foisonnements*

morphologiques : études en hommage à Françoise Kerleroux, pp. 19–44. Presses Universitaires de Paris-Ouest: Nanterre, 2014.
[8] Boyé, Gilles. *Problèmes de morphophonologie verbale en français, espagnol et italien*. PhD dissertation, University Paris 7, 2000.
[9] Cotterell, Ryan and Schütze Hinrich. Joint Semantic Synthesis and Morphological Analysis of the Derived Word. *Transactions of the Association for Computational Linguistics,* 2017.
[10] Fabre, Cécile, Floricic, Franck and Hathout, Nabil. Collecte outillée pour l'analyse des emplois discordants des déverbaux en -eur. Paper presented at the *Journées d'étude sur la place des méthodes quantitatives dans le travail du linguiste*, ERSS, Toulouse, 2004.
[11] Habash, Nizar and Dorr, Bonnie (2003) A categorial variation database for English. In *Proceedings of NAACL 2003*, pp. 96-102.
[12] Hathout, Nabil. Morphonette: a paradigm-based morphological network, pp. 243–262. *Lingue e linguaggio*, 2011(2), 2011.
[13] Hathout, Nabil and Namer, Fiammetta. Démonette, a French derivational morpho-semantic network, pp. 125-168. *LiLT* 11, 2014.
[14] Hathout, Nabil and Namer, Fiammetta. Giving Lexical Resources a Second Life: Démonette, a Multi-sourced Morpho-semantic Network for French. In *Proceedings of LREC 2016*, Portorož, pp. 1084–1091, 2016. European Language Resources Association (ELRA): Slovenia.
[15] Lazaridou, Angeliki, Marelli, Marco, Zamparelli, Roberto and Baroni, Marco. Compositional-ly derived representations of morphologically complex words in distributional semantics. In *Proceedings of ACL 2013,* pp. 1517–1526, Association for Computational Linguistics: Sofia, 2013.
[16] Montermini, Fabio and Boyé, Gilles. Stem relations and inflection class assignment in Italian. *Word Structure 5* (1), pp. 69–87, 2012.
[17] Montermini, Fabio and Bonami, Olivier. Stem spaces and predictability in verbal inflection, *Lingue e linguaggio 2013 (2)*, pp. 171–90, 2013.
[18] Namer, Fiammetta. *Morphologie, Lexique et TAL: l'analyseur DériF* Hermes Sciences Publishing: London, 2009.
[19] Namer, Fiammetta. A Rule-Based Morphosemantic Analyzer for French for a Fine-Grained Semantic Annotation of Texts. in Mahlow Cerstin and Piotrowski, Michael (eds.), *SFCM 2013*, pp. 93–115, Springer: Heidelberg, 2013.
[20] New, Boris, Lexique 3: Une nouvelle base de données lexicales. In *Proceedings of TALN*, Louvain, pp. 892–900, 2006.
[21] Pirrelli, Vito and Battista, Marco. The paradigmatic dimension of stem allomorphy in Italian. *Rivista di linguistica* 12 (2), pp. 307–380, 2000.
[22] Plénat, Marc. Le conditionnement de l'allomorphie radicale en français. *Mémoires de la Société de Linguistique de Paris*, Nouvelle série, n° 17, pp. 119–140, 2009.
[23] Roché, Michel. Base, thème, radical, *Recherches linguistiques de Vincennes*, 39 (1), pp. 95–134, 2010.

[24] Sajous, Franck, Hathout, Nabil and Calderone, Basilio. Glàff, un gros lexique à tout faire du français. In *Actes de la 20e conférence TALN*, pp. 285–298, 2013.
[25] Spencer, Andrew. Identifying stems *Word Structure* 5 (1), pp. 88–108, 2012.
[26] Talamo, Luigi, Celata, Chiara, & Bertinetto, Piermarco. DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure 9* (1), pp: 72–102, 2016.
[27] Tanguy, Ludovic and Hathout, Nabil. Webaffix: un outil d'acquisition morphologique dérivationnelle à partir du Web. In *Proceedings of the 9e Conférence TALN*, pp. 245–254. ATALA: Nancy, 2002.
[28] Tribout, Delphine. Verbal stem space and verb to noun conversion in French. *Word Structure*, 5, pp. 109–28, 2012.
[29] Wells, J. SAMPA computer readable phonetic alphabet. *Handbook of Standards and Resources for Spoken Language Systems*. D. Gibbon, R. Moore and R. Winski. Berlin/New York, Mouton de Gruyter. Part IV, section B, 1997.
[30] Zeller, Britta, Snajder, Jan and Padó, Sebastian. DErivBase: Inducing and Evaluating a Derivational Morphology Resource for German. In *Proceeding of the 51th ACL 2013,* pp. 1201-1211. ACL: Sofia, 2013.