

A Lexical Network with a Morphological Model in It

Nabil Gader

MVS Publishing Solutions
395, chemin de la Cartonnerie
88100 Sainte-Marguerite, France
nabil.gader@mvs.fr

Aurore Koehl

ATILF CNRS
44, avenue de la Libération, BP 30687
54063 Nancy Cedex, France
aurore.koehl@univ-lorraine.fr

Alain Polguère

Université de Lorraine & ATILF CNRS
44, avenue de la Libération, BP 30687
54063 Nancy Cedex, France
alain.polguere@univ-lorraine.fr

Abstract

The French Lexical Network (fr-LN) is a global model of the French lexicon presently under construction. The fr-LN accounts for lexical knowledge as a lexical network structured by paradigmatic and syntagmatic relations holding between lexical units. This paper describes how morphological knowledge is presently being introduced into the fr-LN through the implementation and lexicographic exploitation of a dynamic morphological model. Section 1 presents theoretical and practical justifications for the approach which we believe allows for a cognitively sound description of morphological data within semantically-oriented lexical databases. Section 2 gives an overview of the structure of the dynamic morphological model, which is constructed through two complementary processes: a Morphological Process—section 3—and a Lexicographic Process—section 4.

1 Introduction

We present a morphological model implemented in order to feed the French Lexical Network database—hereafter fr-LN—, presently under development at the ATILF CNRS lab. The fr-LN belongs to the broad family of lexical resources designed as networks of lexical units (Fellbaum, 1998; Baker *et al.*, 2003; Ruppenhofer *et al.*, 2010; Spohr, 2012). Its design, content and mode of construction has already been documented in various publications (Lux-Pogodalla and Polguère, 2011; Gader *et al.*, 2012; Polguère, 2014; Polguère, to appear) and we strictly focus here on its newly developed morphological component.

The morphological description of French lexemes discussed below possesses two main characteristics:

1. it is dynamically created from a full-fledged grammatical model of French inflectional morphology;
2. it is meant to be used in the context of a lexicographic project where morphological tables are individually associated to senses of (polysemic) vocables, thus accounting for potential morphological discrepancies between senses within a given vocable.

We believe that our approach allows for a cognitively sound implementation of morphology in lexical databases that are primarily oriented towards the description of senses (rather than forms). Indeed, we do not simply inject lists of lexical forms into the fr-LN database but describe *morphological knowledge* by means of a “true” model of French inflectional morphology.

This work is licenced under a Creative Commons Attribution 4.0 International License. Page numbers and proceedings footer are added by the organizers. License details: <http://creativecommons.org/licenses/by/4.0/>

1.1 Interconnection between lexical and morphological models

There are at least two main characteristics that a cognitively sound lexical model of a given language has to possess.

- First and foremost, it must account for the Speaker’s knowledge of lexical rules (a lexical unit and all its properties being considered here as being a cluster of lexical rules).
- Second, it must be structured in a way that makes it possible to support the modeling of (i) “natural” processes of lexical knowledge evolution—acquisition, consolidation, relativization, loss of lexical knowledge—and (ii) linguistic processes of speech, understanding, paraphrase, translation, word association (Dunbar, 2012), etc.

Lexical knowledge, however, is intricately related to grammatical knowledge, to the point that it is not necessarily possible to radically separate lexical information from grammatical information in a cognitive lexicon. This is particularly true for language morphology, that can be said to belong simultaneously to both the lexical and grammatical modules of natural languages. Consequently, a lexical model that aims at cognitive relevance has to take the morphological bull by the horns and handle morphological properties and behavior of lexical units by modeling actual morphological knowledge.

Not all lexical resources link a lexical unit to its forms and morphology is often “externalized,” using separate dedicated resources (see section 1.2 below). However, for a language that is rather rich in terms of inflectional morphology, such as French, it is particularly interesting to embed in lexical resources an explicit morphological model. The purpose of such model is to allow lexicographers to account for inflected forms of lexical units by associating each individual unit (= sense) to inflectional classes and dynamically obtain in the process morphological tables of all corresponding lexical forms.

Before we proceed, let us enumerate the basic terminology that will be used throughout the paper.

- *Lexical units* are of two kinds: (i) *lexemes*—CHEVAL I.1a ‘horse’ ~ CHEVAL I.1b ‘horse riding’—are monolexicemic lexical units; (ii) *idioms*—CULOTTE DE CHEVAL ‘saddlebags’—are syntagmatic lexical units.
- *Vocables*—CHEVAL—are (potentially) polysemic words. They are modeled in the fr-LN as sets of lexical units connected by a relation of copolysemy.
- *Wordforms* are linguistic signs expressing lexemes—singular *cheval* I.1a and plural *chevaux* I.1a are the two wordforms for the CHEVAL I.1a lexeme. Because wordforms are linguistic signs, each individual wordform has to be described as a <signified, signifier, restricted combinatorics> triplet (Mel’čuk, 2012, Chapter 1).

1.2 Current approaches to morphology in sense-oriented lexical databases

In this section, we briefly summarize the treatment of morphological information in major sense-oriented lexical databases and explain why we decided to elaborate an approach of our own.

Our initial constraints were that we wanted to truly handle forms related to lexemes with respect to general rules of inflectional morphology. Additionally, we wanted to model in an elegant way phenomena such as spelling variation (*cuillère* ‘spoon’ ~ *cuiller*), euphony (*j’aime* ‘I love’ ~ *aimé-je*), alternative inflected forms (*je m’assois* ~ *je m’assieds*) or defectiveness (*je fris du lard* ‘I fry bacon’, but there is no corresponding 1st person plural; one has to say *nous faisons frire du lard* lit. ‘We make bacon fry’).

To our knowledge, no current general purpose lexical database—for French or other languages—currently meet these requirements. WordNet, for instance, only stores base forms of lexemes and has no embedded morphological model of English. An external lemmatizer, Morphy,¹ is used to access lexical senses via inflected forms. The situation is different in FrameNet. As indicated in (Ruppenhofer *et al.*, 2010, p. 93–94), lemmas are stored in the *Lexical Database* component of FrameNet, together with corresponding wordforms. However, no grammatical model of inflection is embedded in the database and made available for lexicographic purposes.

¹<https://wordnet.princeton.edu/wordnet/man/morphy.7WN.html>

2 Dynamic approach to morphological description

This section is devoted to the presentation of the morphological model embedded in the fr-LN. We proceed in two steps. Firstly (section 2.1), we detail the limitations of existing morphological databases for French, which explain why we decided to not “inject” their content in the fr-LN. Secondly (section 2.2), we present the general design of our morphological model and detail its dynamic nature.

2.1 Limitations of existing morphological resources for French

In order to model “morphological knowledge” within a lexical database, one can either make use of an already existing morphological resource (that will be connected to or embedded into the database), or develop a specific, tailor-made morphological database module—see (Issac, 2010) for a detailed discussion. There exist indeed several morphological resources for French that, in principle, could have been used as embedded morphological modules in the fr-LN. We will explain why limitations found in these resources have led us to choose the second option and design our own morphological model.

We have mainly examined six morphological resources for French, all developed during the past ten years: *Manulex*, *Morphalou*, *Lexique 3*, *Lefff*, *Flexique* and *Morfetik*.² Here is a brief recap of the observations we have made, based on our specific needs and expectations. For lack of space, we cannot make a detailed presentation of these resources and our evaluation will by necessity be rather sketchy.

Manulex was designed for psycholinguistic research (Lété *et al.*, 2004). It contains 48,886 French wordforms. The list of wordforms results from a “grade-based word frequency list extracted from a corpus of first to fifth grade readers used in French elementary schools” (Lété *et al.*, 2004, p. 159). *Manulex* has therefore a limited coverage, when compared to other existing resources that target the bulk of the French lexicon and can store up to 500,000 forms.

But coverage is not the only issue. The quality of data can vary greatly from database to database. In *Morphalou* (Romary *et al.*, 2004), for instance, one can find a lot of miscategorizations and misspellings. Reusing *Morphalou*’s data would thus raise many maintenance issues.

While having a larger coverage than *Manulex* and data of better quality than *Morphalou*, *Lexique 3* (New, 2006) poses several problems of its own. First, inflectional paradigms are not complete, because *Lexique 3*’s wordlist was extracted from the Frantext corpus (Montémont, 2008), that contains only part of the lexicon of contemporary French. Second, pairs like *chat* ‘cat’ ~ *chatte* ‘female cat’ have been encoded as one entry, which contradicts our theoretical and descriptive choices. Following (Mel’čuk, 2000; Delaite and Polguère, 2013), we consider that no inflectional mechanism is involved here. There are two distinct CHAT ~ CHATTE nominal lexemes in French; the feminine is morphologically **derived** (i.e. produced by morphological derivation) from the masculine and has to be accounted for separately. Both aspects—incompleteness and inappropriate descriptive postulates such as in the case of $N_{\text{masc}} \sim N_{\text{fem}}$ pairs—disqualified *Lexique 3* in our quest for an already-existing resource.

Flexique (Bonami *et al.*, 2013) is derived from *Lexique 3*, but the problems we just mentioned are solved: paradigms are now complete and pairs like CHAT ~ CHATTE have been encoded as two separate entries. However, *Flexique*—just as the two remaining resources *Lefff*³ (Sagot, 2010) and *Morfetik* (Mathieu-Colas, 2009)—lacks alternative forms for inflections⁴ or orthographic variants, as reported by the authors themselves.

Last but not least, all resources cited above associate a morphological description to a lexical entry, **not to a specific sense**. However, not all senses of a given polysemic vocable necessarily possess the same morphological behavior, and this is valid for most of natural languages. For instance, the sense ‘flag or other symbolic object’ of the COLOR vocable is plural only (*to raise the colors*). In other words:

²These resources are all available for research. The dictionaries of the *Antidote* suite incorporate a powerful morphological model for French (Antidote, 2014). However, *Antidote* is a commercial product; we cannot examine its internal design and its morphological model is of course not available for embedding in a lexical database such as fr-LN, whose linguistic content will be freely available.

³*Lexique des Formes Fléchies du Français*.

⁴For instance, these resources do not indicate that Fr. AIL ‘garlic’ has two alternative plural forms *ails* ~ *aulx* (section 4.1 below) or that S’ASSEOIR ‘to sit’ has two alternative forms for most of its inflections—*je m’assoie* ~ *je m’assieds*, *tu t’assois* ~ *tu t’assieds*, *il s’assoit* ~ *il s’assied*...

A morphological model encapsulated in a lexical database should describe actual wordforms: linguistic signs made up of a signified, a signifier and combinatorial properties. Signifiers should not remain disconnected from the signified they express.

In this respect, and to our knowledge, there exists no sense-based morphological model available for French prior to our work. This left us with no choice but to design a model of our own, that would be specially designed to accompany our lexicographic project and be better suited for applications such as word sense identification backed by lexical knowledge (Leacock and Chodorow, 1998).

2.2 General design of the fr-LN dynamic morphological model

The core of our morphological model is a set of *Morphological Templates* that define corresponding inflection classes as *Prototypical Tables* of inflection. These latter tables are named after a lexeme that prototypically represents the corresponding morphological paradigm: Prototypical Table of nouns of the CHAT ‘cat’ family, of verbs of the DANSER ‘dance’ family, etc. The association of a Prototypical Table to a given lexeme automatically generates one or more *Lexeme Table(s)*, i.e. tables that contain the description of all wordforms expressing this lexeme. Wordforms themselves are defined as relations holding between three database elements: (i) a given Lexeme Table, (ii) a set of grammatical features (mainly, grammemes) associated to the wordform and (iii) a given signifier.

The integration of morphological knowledge into the fr-LN database is performed through two complementary processes, as visualized in Figure 1:

1. a *Morphological Process*—construction of Morphological Templates from which Prototypical Tables are generated;
2. a *Lexicographic Process*—creation of Lexeme↔Prototypical Table(s) associations, from which Lexeme Tables are automatically derived.

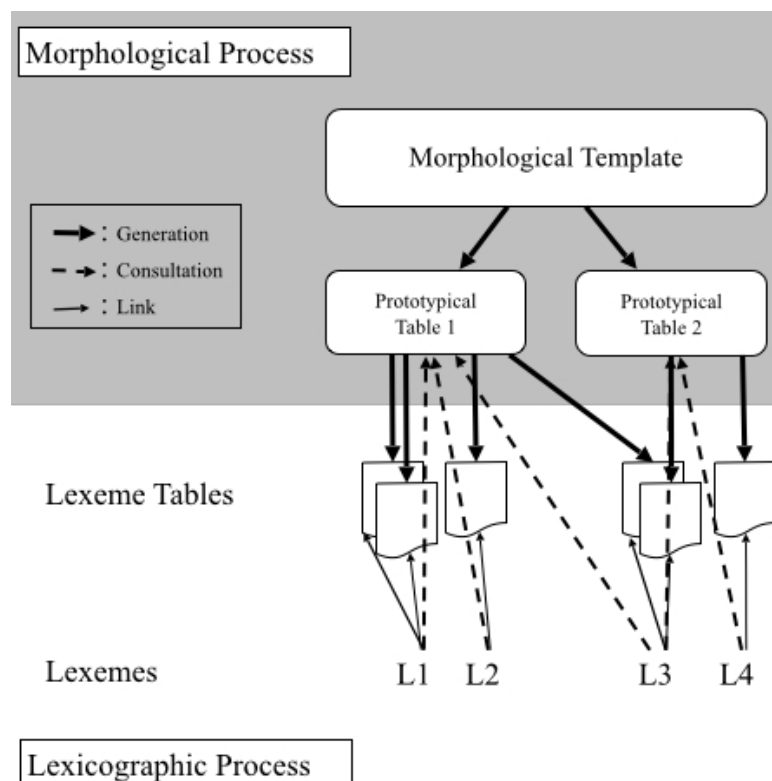


Figure 1: The fr-LN dynamic morphological model

Morphological Templates and Prototypical Tables are the core modules of our dynamic morphological model. In section 3 below, we describe the Morphological Process that leads to the creation of Morphological and Prototypical Tables. We then proceed, in section 4, with the Lexicographic Process that leads to the generation of individual Lexeme Tables. In both sections, we use examples to illustrate the descriptive power and flexibility of the approach in the context of our lexicographic enterprise.

3 Morphological Process → Morphological Templates and Prototypical Tables

The construction of the morphological model—Morphological and Lexicographic Processes—is performed with the *Dicet* editor (Gader *et al.*, 2012), the same lexicographic editor used to build the fr-LN lexical graph through weaving of lexical relations. Illustrative figures in this section and the next one are screen dumps of access to the morphological model by means of *Dicet*.

3.1 Morphological Templates

The role of Morphological Templates is to establish parameter and variable slots that are common to sets of related Prototypical Tables. For instance, the *Adjectifs* ‘Adjectives’ template, shown in Figure 2, is used to generate all adjectival Prototypical Tables.

Identifier : 12

Name : Adjectifs

Parameters :

	Name	Occurrences			
1	Base	1			
2	Variation(s)	0..n			

Variables :

	Name	Occurrences			
1	Genre	1	↺	↻	✖
2	Nombre	1			
3	Tronquer	0..1			
4	Ajouter	0..1			
5	Suffixe	0..1			
6	Variation(s)	0..n			

Compatible parts of speech

- ▶ MARQUES D'USAGE
- ▶ CARACTÉRISTIQUES GRAMMATICALES FO
- ▶ FLEXION ET AUTRES CARACTÉRISTIQUES
- ▶ POSITION SYNTAXIQUE

- ▼ CARACTÉRISTIQUES GRAMMATICALES FO
 - ▼ Partie du discours
 - ▼ Adj
 - adjectif

Figure 2: The *Adjectifs* Morphological Template

Figure 2 indicates that all Prototypical Tables that are created from the *Adjectifs* template will have the same set of columns, defined in the table called *Variables*: gender, number, truncation performed on the string of characters that corresponds to the stem, addition to it, suffixation and variation(s).

More generally, variables correspond to either:

- grammemes expressed by the wordforms;
- formal adjustments to be performed on the stem (truncation from/addition to the stem, suffixation);
- possible variations of given wordforms in the table.

Each wordform is related to a particular set of variables. For example, the wordform *actives*—which means ‘active’ ⊕ feminine ⊕ plural⁵ and whose stem is *actif*—is associated to the following set of vari-

⁵The ⊕ operator represents the linguistic union of (components of) two linguistic signs (Mel’čuk, 2012, Chapter 1).

able instantiations: Genre=feminine, Nombre=plural, Tronquer=1, Ajouter=-v-, Suffixe=-es, Variation(s)=∅.

The Parameters table specifies the characteristics that are shared by all tables of a given template. For instance, all adjectival tables contain a Base ‘stem’ field and a Variation(s) field. This means that an adjectival table will force the lexicographer to declare the stem used to generate all the wordforms of the adjective and will allow her to declare possible variants for all wordforms of the table, which will condition the generation of more than one Lexeme Table.

Let us illustrate this with the lexeme ABÎMÉ ‘damaged’. Declaring the morphology of this lexeme will trigger the generation of two morphological tables: one for the (default) “traditional” spelling (1.) and one for the “rectified” spelling (2.).⁶

1. ABÎMÉ {*abîmé*_(masc, sing), *abîmés*_(masc, plur), *abîmée*_(fem, sing), *abîmées*_(fem, plur)};
2. ABÎMÉ {*abimé*_(masc, sing), *abimés*_(masc, plur), *abimée*_(fem, sing), *abimées*_(fem, plur)}
Variation(s)=rectified spelling.

Because of the rather rich morphology of French verbs, verbal templates require more parameters and variables than nominal or adjectival ones. For instance, verbal templates require two additional parameters in order to deal with (i) choice of auxiliary for compound tenses and (ii) possible use of the SE reflexive pronoun if the verb is pronominal (e.g. SE SUICIDER ‘to commit suicide’)—see section 4.3.

Notice that the rationale behind the use of Morphological Templates is the need to design a generic approach that will allow us to work on typologically unrelated languages—cf. final remarks in section 5. It is at the level of Morphological Templates that general principles of word construction are encoded, for each individual language.

3.2 Prototypical Tables

As mentioned earlier, Prototypical Tables are generated from Morphological Templates: they feature actualizations of all characteristics (parameters and variables) defined in their source template. **In other words, Prototypical Tables represent morphological classes.** At the time of writing, 15 Prototypical Tables have been created using the Noms ‘Nouns’ template, 34 using Adjectifs and almost a hundred Prototypical Tables have been created to account for French verbal morphology.

Figure 3 illustrates the approach with the petit ‘small’ Prototypical Table. This table instantiates the Adjectifs Morphological Template with Base=*petit* and Variation(s)=∅ as parameters.

Identifier :		Name :		Template :							
16		petit		Adjectifs							
Parameters :											
Base				Variation(s)							
petit											
Rules :											
	N...	Genre	Nombre	Tronquer	Ajouter	Suffixe	Variatio...	Form			
2	<input checked="" type="checkbox"/>	Masculin	Singulier					petit			
3	<input type="checkbox"/>	Masculin	Pluriel			s		petits			
4	<input type="checkbox"/>	Féminin	Singulier			e		petite			
5	<input type="checkbox"/>	Féminin	Pluriel			es		petites			

Figure 3: The petit ‘small’ Prototypical Table

⁶The French language council of France—*Conseil de la langue française*—has officially introduced a new spelling system in 1990 (Conseil supérieur de la langue française, 1990), which concerns around 5,000 words and whose usage has been declared to be facultative. Some 25 years later, both the traditional and rectified systems are still cohabiting, even in official texts and at school, which pretty much disqualifies the spelling reform as being a stunning success. Based on current trends (or lack thereof), the approach of the fr-LN database is to maintain the two systems, using the traditional spelling as default one.

In Figure 3, the `Rules` table displays the result of the creation of the Prototypical Table, where each line corresponds to a given dynamically created wordform of `PETIT`. In each individual “rule” (i.e. grammatical characterization of the corresponding wordform) the gender and number features are instantiated with one grammeme of the pairs masculine ~ feminine and singular ~ plural, respectively. The rule corresponding to the wordform that functions as lexicographic name for the lexeme is checked on the left-hand side. (The lexicographic name of a lexeme is the default form that will be used to name this lexeme in dictionary wordlists, articles, etc.)

Remember that the `petit` Prototypical Table (Figure 3) is distinct from the Lexeme Table of `PETIT`: this latter is the model of an adjectival **inflectional class** identified as being, prototypically, that of `PETIT`. Actual Lexeme Table are produced in the context of a Lexicographic Process, that is our next topic.

4 Lexicographic Process → Lexeme Tables

The Lexicographic Process is operated through the creation of an association between given lexemes and given Prototypical Tables. Each association performed on a lexeme by the lexicographer produces the generation of one or more Lexeme Table(s) for this lexeme. We detail this process successively for nouns (4.1), adjectives (4.2) and verbs (4.3).

4.1 Nominal Lexeme Tables

French nouns carry grammatical number—singular or plural. The singular is expressed by a \emptyset -suffix (no addition to the stem). Canonical nominal plural in French is formed by suffixing the *-s* suffix to the nominal stem. As an example, the two wordforms of `ACTEUR` ‘actor’ are singular *acteur* and plural *acteurs*. All nominal lexemes inflecting in the canonical way are associated with the Prototypical Table `chat`. Figure 4 illustrates how this association is performed with the `Dicet` editor.

Identifiers	Inflection table	Naming form	Probability
632/9	chat	acteur	100

Figure 4: `ACTEUR` lexeme ↔ `chat` Prototypical Table association

There are however cases where the nominal plural is not formed by suffixing *-s*. Four cases can be mentioned.

1. There are unmarked nominal plurals. This concerns nouns ending with *-s*, *-z* or *-x*, like `ABUS` ‘[an] abuse’, `PRIX` ‘price’ or `RIZ` ‘rice’, which are invariable. The Prototypical Table `nez` ‘nose’ has been created to handle such cases. The association of the Prototypical Table `nez` with a nominal stem generates two wordforms (one for singular and one for plural) which have an identical signifier (namely, that of the stem).
2. Some nouns are irregular: their inflected forms cannot be computed by means of general morphological rules. For instance, the plural form of `AIL` ‘garlic’ is *aulx*, though it can also be expressed by the regular form *ails*. To account for this, the lexeme `AIL` has been connected to the special Prototypical Table `ail` which generates a Lexeme Table containing both the regular and irregular plurals: $\{ail_{(sing)}, ails_{(plur)}, aulx_{(plur)}\}$.
3. Lexemes can be defective: there is an “empty cell” (or more) in their table of wordforms (Baerman *et al.*, 2010). For nouns, the defective form is of course either the singular or the plural:
 - defective singular: `COULEURS III.1b` ‘colors (= flag or other symbolic object)’;
 - defective plural: `CIGARETTE II` ‘habit of smoking’ (*Je devrais arrêter la cigarette* ‘I should quit smoking cigarette’).

4. Some nouns can have spelling variants for their stem. (Catach, 1995) and (Sebba, 2003) studied spelling variations concerning the use of accents, the hyphen, archaic forms, the plural of compound words and double consonants. These possible orthographic variants are recorded as spelling variants in our resource. For example, the noun CUIILLÈRE ‘spoon’ has two spellings, *cuillère* and *cuiller*. Consequently, the lexeme CUIILLÈRE has been coupled with two tables, generating respectively the wordforms {*cuillère*_(sing), *cuillères*_(plur)} and {*cuiller*_(sing), *cuillers*_(plur)}. Most of the time, spelling variation of the stem follows the last orthographic reform of 1990 (footnote 6 above); in this case, the form recommended by the reform is labelled as “rectified spelling.”

Unmarkedness (ABUS), irregularity (AIL), defectiveness (COULEURS III.1b) and spelling variation (CUIILLÈRE) concern not only nouns but also lexemes of other parts of speech, as will be seen in the next sections.

4.2 Adjectival Lexeme Tables

As shown earlier in section 3.1, French adjectives carry both grammatical gender (masculine ~ feminine) and number (singular ~ plural). A few associations of Prototypical Tables with adjectival lexemes are used as illustrations in what follows.

Canonical inflection of French adjectives—namely, feminine formed by *-e* suffixation and plural by *-s* suffixation—is modeled in the *petit* ‘small’ Prototypical Table. The association of this table with an adjectival lexeme such as ABSENT_{Adj} ‘absent’ dynamically generates the table of all corresponding wordforms: {*absent*_(masc, sing), *absents*_(masc, plur), *absente*_(fem, sing), *absentes*_(fem, plur)}.

Additionally, a significant number of French adjectives are unmarked for gender—e.g. EFFICACE ‘efficient’ {*efficace*_(masc, sing), *efficaces*_(masc, plur), *efficace*_(fem, sing), *efficaces*_(fem, plur)}. Their wordforms are generated using the *aimable* Prototypical Table.

Beside the two above-mentioned regular cases, many adjectives have rather idiosyncratic behavior. This includes invariability, allomorph stems or spelling variations.

1. Invariable adjectives are lexemes whose wordforms (inflected forms) are based on the same signifier. For example, the adjective DEBOUT ‘standing up’ possesses the formally identical wordforms {*debout*_(masc, sing), *debout*_(masc, plur), *debout*_(fem, sing), *debout*_(fem, plur)}, that are generated using the *carmin* ‘of carmine color’ Prototypical Table.
2. Stem allomorphy can be exemplified with SEC_{Adj} ‘dry’ {*sec*_(masc, sing), *secs*_(masc, plur), *sèche*_(fem, sing), *sèches*_(fem, plur)} OR BREF_{Adj} ‘brief’ {*bref*_(masc, sing), *brefs*_(masc, plur), *brève*_(fem, sing), *brèves*_(fem, plur)}. It is dealt with on a case-by-case basis, with the generation of specific Lexeme Tables.
3. As for nouns, we have to deal with spelling variation of adjectival stems—see the case of ABÎMÉ ‘damaged’ mentioned in section 3.1 above. This implies the creation of two (or more) Lexeme Tables for the same lexeme, one for each possible stem.

Another difficulty we had to deal with comes from the fact that adjectives may have a particular form when they are linearized before a vowel-initial noun (Bonami and Boyé, 2005). Such is the case of VIEUX ‘old’:

- (1) a. *Ugo, c’était un vieux copain d’enfance.*
‘Ugo was an old childhood friend’
[Frantext, IZZO Jean-Claude, *Total Khéops*, 1995, p. 41]
- b. *Après tout, je suis ton plus vieil ami.*
‘After all, I’m your oldest friend’
[Frantext, BEAUVOIR (de) Simone, *Les Mandarins*, 1954, p. 364]

In order to handle such cases, the lexeme VIEUX_{Adj} is related to the **five** rather than four wordforms in its Lexeme Table, *vieil* being encoded as a variant wordform for masculine singular—see Figure 5.

Modèles de flexion :		Table de flexion :						
Adjectifs		vieux						
Paramètres :								
Base		Variation(s)						
vieux								
Règles :								
	Genre	Nombre	Tronquer	Ajouter	Suffixe	Variation(s)	Forme	Fo...
2	Masculin	Singulier					vieux	<input type="checkbox"/>
3	Masculin	Singulier	2	il		variation	vieil	<input type="checkbox"/>
4	Masculin	Pluriel					vieux	<input type="checkbox"/>
5	Féminin	Singulier	2	ill	e		vieille	<input type="checkbox"/>
6	Féminin	Pluriel	2	ill	es		vieilles	<input type="checkbox"/>

Figure 5: Lexeme Table of VIEUX ‘old’

Finally, we have included in adjectival Lexeme Tables wordforms that are **linguistically** possible and attested, though they may seem deviant for **conceptual** reasons. For instance, ENCEINTE_{Adj} ‘pregnant’ is naturally related to two feminine wordforms: *enceinte*_(fem, sing) ~ *enceintes*_(fem, plur); but in the eventuality that one does want to talk about a pregnant man (for instance, in order to state that this would be a challenging situation), two distinct pairs of masculine wordforms can be used *enceinte*_(masc, sing) ~ *enceintes*_(masc, plur) or *enceint*_(masc, sing) ~ *enceints*_(masc, plur). See the following examples found on the Internet:

- (2) a. *Des jeunes garçons enceintes, c’est ce que voient les habitants de Chicago sur leurs panneaux publicitaires.*
‘Young pregnant boys, that’s what people in Chicago see on advertisement billboards’
[<http://www.grazia.fr/societe/news/etats-unis-des-hommes-enceintes-pour-promouvoir-la-contraception-551492>]
- b. *À Chicago, les affiches publicitaires mettant en scène de jeunes garçons enceints ont remplacé celles, plus classiques, sur les préservatifs et la pilule.*
‘In Chicago, advertisements showing pregnant teenage boys have replaced more traditional ones, about condoms and the birth control pill’
[<http://www.terrafemina.com/vie-privee/sexo/articles/27026-contraception-des-garcons-enceints-pour-sensibiliser-les-ados-de-chicago.html>]

These forms, that are amply attested, are labelled as possible in the Lexeme Table of ENCEINTE_{Adj}, where the feminine singular wordform is of course identified as naming form; cf. Figure 6.

Modèles de flexion :		Table de flexion :						
Adjectifs		enceinte						
Paramètres :								
Base		Variation(s)						
enceinte								
Règles :								
	Genre	Nombre	Tronquer	Ajouter	Suffixe	Variation(s)	Forme	Fo...
2	Féminin	Singulier					enceinte	<input checked="" type="checkbox"/>
3	Féminin	Pluriel			s		enceintes	<input type="checkbox"/>
4	Masculin	Singulier	1			possible	enceint	<input type="checkbox"/>
5	Masculin	Pluriel	1		s	possible	enceints	<input type="checkbox"/>
6	Masculin	Singulier				possible	enceinte	<input type="checkbox"/>
7	Masculin	Pluriel			s	possible	enceintes	<input type="checkbox"/>

Figure 6: Lexeme Table of ENCEINTE_{Adj} ‘pregnant’

This approach reflects actual usage and is more valid from a linguistic point of view than the alternative solution that consists in encoding ENCEINTE_{Adj} as defective adjective.

4.3 Verbal Lexeme Tables

For lack of space, we provide only an outline of how the rich inflectional morphology of French verbs is being handled in our model. We focus on the most significant aspects of the question only.

The pairing of a Prototypical Table with a verbal lexeme implies that information is provided on (i) which auxiliary (*avoir* or *être*) is selected by the verb for compound tenses and (ii) whether the verb is pronominal (*S'AMÉLIORER* ‘to become better’) or not (*AMÉLIORER* ‘to improve (something)’). This is illustrated in Figure 7, which shows a short sample of the Lexeme Table of the verb *AGACER I* ‘to annoy’.

Modèles de flexion :		Table de flexion :									
Verbes du premier groupe		avancer									
Paramètres :											
Base	Pronominal	Auxiliaire(s)	Variation(s)								
agac		avoir									
Règles :											
	Mode	Temps	Genre	Personne	Nombre	Tronquer	Ajouter	Suffixe	Variation(s)	Forme	Fo...
2	Indicatif	Présent		Premièr...	Singulier			e		j'agace	<input type="checkbox"/>
3	Indicatif	Présent		Premièr...	Singulier			é	eupho_...	agacé-je	<input type="checkbox"/>
4	Indicatif	Présent		Deuxiè...	Singulier			es		tu agaces	<input type="checkbox"/>
5	Indicatif	Présent		Troisiè...	Singulier			e		il, elle	<input type="checkbox"/>
6	Indicatif	Présent		Premièr...	Pluriel	1	ç	ons		nous	<input type="checkbox"/>
7	Indicatif	Présent		Deuxiè...	Pluriel			ez		vous	<input type="checkbox"/>
8	Indicatif	Présent		Troisiè...	Pluriel			ent		ils, elles	<input type="checkbox"/>
9	Indicatif	Imparfait		Premièr...	Singulier	1	ç	ais		j'agaçais	<input type="checkbox"/>
10	Indicatif	Imparfait		Deuxiè...	Singulier	1	ç	ais		tu	<input type="checkbox"/>
11	Indicatif	Imparfait		Troisiè...	Singulier	1	ç	ait		il, elle	<input type="checkbox"/>
12	Indicatif	Imparfait		Premièr...	Pluriel			ions		nous	<input type="checkbox"/>
13	Indicatif	Imparfait		Deuxiè...	Pluriel			iez		vous	<input type="checkbox"/>

Figure 7: Lexeme Table (sample) of *AGACER I* ‘to annoy’

At present, 34 different Prototypical Tables—such as *danser* ‘to dance’—have been constructed in order to generate Lexeme Tables for French verbs of the first conjugation class (*premier groupe*, in French grammatical terminology), i.e. verbs that take the *-er* infinitive suffix. Most of these tables were created in order to handle stem alternations, such as the alternation *agac-* ~ *agaç-* for *AGACER I* in Figure 7.

The first conjugation class has the highest cardinality and it is basically the only productive one in contemporary French. (Neologisms normally belong to this class.) There are two other conjugation classes. Verbs of the second conjugation class are dealt with using 3 Prototypical Tables, and 52 Prototypical Tables have been constructed for verbs of the third class.

As for nouns and adjectives, verbs can be defective—11 Prototypical Tables handle defective paradigms—and can have spelling variants. In addition, the morphological model has to deal with suppletive verbs—on suppletion, see (Mel’čuk, 1994; Bonami and Boyé, 2003; Corbett, 2007).

To conclude our description of the morphological Lexicographic Process, it is important to mention the fact that an inheritance mechanism has been implemented in the Dicet editor. Senses that are created inside an already existing vocable automatically inherit their morphological description from the basic lexical unit of the vocable (the sense controlling the vocable’s polysemic structure). Inherited morphological data get a default measure of confidence of 50%. It has to later be either validated by lexicographers (measure of confidence pushed to 100%) or manually overwritten, if the sense in question has a specific morphological behavior.

5 Concluding remarks

The production of individual Lexeme Tables has started only four months ago, after the complex tasks of designing and implementing the dynamic morphological model had been completed. For the time

being, approximately 10% of the vocables (= entries) currently present in the database have been morphologically described.⁷

We expect to have finished the morphological work “on back order” in the fr-LN database within a few months. Future developments include:

- the treatment of compounds such as BOULANGER-PÂTISSIER lit. ‘baker-pastry maker’ {*boulangier-pâtissier*_(sing), *boulangers-pâtissiers*_(plur)} (Mathieu-Colas, 2011);
- the computation of inflected forms of idioms using the encoding of their lexico-syntactic structure;
- the application of our dynamic approach to the modeling of morphology of languages other than French within their lexical networks, starting with the English Lexical Network (en-LN) presented in (Gader *et al.*, 2014).

Acknowledgments

Work on the fr-LN is supported by a grant from the Agence de Mobilisation Économique de Lorraine (AMEL) and Fonds Européen de Développement Régional (FEDER). We thank Cogalex 2014 anonymous reviewers for their comments on a first version of this paper.

References

- Antidote. 2014. *Antidote 8* [software]. Druide informatique inc., Montreal, QC.
- Matthews Baerman, Greville G. Corbett and Dunstan Brown. 2010. *Defective Paradigms: missing forms and what they tell us*. Oxford University Press, Oxford, UK.
- Collin F. Baker, Charles J. Fillmore and Beau Cronin. 2003. The Structure of the FrameNet Database. *International Journal of Lexicography*, 16(3):281–296.
- Olivier Bonami and Gilles Boyé. 2003. Supplétion et classes flexionnelles dans la conjugaison du français. *Langages*, 37(152):102–126.
- Olivier Bonami and Gilles Boyé. 2005. Construire le paradigme d’un adjectif. *Recherches linguistiques de Vincennes*, 34:77–98.
- Olivier Bonami, Gauthier Caron and Clément Plancq. 2013. Flexique: an inflectional lexicon for spoken French. Technical documentation [<http://www.llf.cnrs.fr/flexique/documentation.pdf>].
- Nina Catach. 1995. Le problème des variantes graphiques: variantes du passé, du présent et de l’avenir. *Langue française*, 25–32.
- Conseil supérieur de la langue française. 1990. *Les rectifications de l’orthographe*. Journal officiel de la République française N° 100. Édition des documents administratifs. Direction des journaux officiels, Paris.
- Greville G. Corbett. 2007. Canonical Typology, Suppletion, and Possible Words. *Language*, 8–42.
- Candice Delaite and Alain Polguère. 2013. Sex-Based Nominal Pairs in the French Lexical Network: It’s Not What You Think. *Proceedings of the 6th International Conference on Meaning-Text Theory (MITT’13)*, Valentina Apresjan, Boris Iomdin, Ekaterina Ageeva (Eds.), Prague, 29–40.
- George Dunbar. 2012. Adaptive Resonance Theory as a model of polysemy and vagueness in the cognitive lexicon. *Cognitive Linguistics*, 23(3):507–537.
- Christiane Fellbaum (Ed.). 1998. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA.
- Nabil Gader, Veronika Lux-Pogodalla and Alain Polguère. 2012. Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CoGALex III)*, The COLING 2012 Organizing Committee, Mumbai, 109–125.

⁷More than 5% of these vocables required separate morphological characterization for their senses (cf. the case of COULEURS III.1b mentioned in section 4.1). This is a rather high proportion—considering the fact that many of the vocables we treated are monosemic and are thus not concerned by potential morphological discrepancies between senses—and it is a strong argument in favor of the design of our sense- rather than vocable-based morphological model.

- Nabil Gader, Sandrine Ollinger and Alain Polguère. 2014. One Lexicon, Two Structures: So What Gives?. *Proceedings of the Seventh Global Wordnet Conference (GWC2014)*, Heili Orav, Christiane Fellbaum, Piek Vossen (Eds.), Global WordNet Association, Tartu, 163–171.
- Fabrice Issac. 2010. A framework for representing lexical resources. *Proceedings of the 23rd International Conference on Computational Linguistics*, Chu-Ren Huang, Dan Jurafsky (Eds.), Beijing, China, 490–497.
- Claudia Leacock and Martin Chodorow. 1998. Combining Local Context with WordNet Similarity for Word Sense Identification. *WordNet: An Electronic Database*, Christiane Fellbaum (Ed.), The MIT Press, Cambridge, MA, 265–283.
- Bernard Lété, Liliane Sprenger-Charolles and Pascale Colé. 2004. MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments and Computers*, 36(1):156–166.
- Veronika Lux-Pogodalla and Alain Polguère. 2011. Construction of a French Lexical Network: Methodological Issues. *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop*, Ljubljana, 2011, 54–61.
- Michel Mathieu-Colas. 2009. Morfetik: une ressource lexicale pour le TAL. *Cahiers de Lexicologie*, 94:137–146.
- Michel Mathieu-Colas. 2011. Flexion des noms et des adjectifs composés: Principes de codage. Technical documentation [<http://halshs.archives-ouvertes.fr/halshs-00635018/>].
- Igor Mel'čuk. 1994. Suppletion: toward a logical analysis of the concept. *Studies in Language*, 18(2):339–410.
- Igor Mel'čuk. 2000. Un FOU/une FOLLE: un lexème ou deux? *Bulag*, hors-série:95–106.
- Igor Mel'čuk. 2012. *Semantics: From meaning to text* (vol. 1). Studies in Language Companion Series 129, John Benjamins, Amsterdam/Philadelphia.
- Véronique Montémont. 2008. Discovering Frantext. *New Beginnings in Literary Studies*, Jan Auracher, Willie van Peer (Eds.), Cambridge Scholars Publishing, Newcastle, UK, 89–107.
- Boris New. 2006. Lexique 3: Une nouvelle base de données lexicales. *Proceedings of TALN 2006*, Chu-Ren Huang, Dan Jurafsky (Eds.), Leuven, Belgium, 490–497.
- Alain Polguère. 2014. Principes de modélisation systémique des réseaux lexicaux. *Proceedings of TALN 2014*, Brigitte Bigi (Ed.), Marseille, France, 79–90.
- Alain Polguère. To appear. From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*.
- Laurent Romary, Susanne Salmon-Alt and Gil Francopoulo. 2004. Standards going concrete: from LMF to Morphalou. *Workshop on Electronic Dictionaries, Coling 2004*, Geneva [<http://hal.inria.fr/docs/00/12/14/89/PDF/LRSSAGFFinal.pdf>].
- Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson and Jan Scheffczyk. 2010. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, CA.
- Benoît Sagot. 2010. The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)*, Istanbul.
- Mark Sebba. 2003. Spelling rebellion. *Pragmatics and beyond*, 151–172.
- Dennis Spohr. 2012. *Towards a Multifunctional Lexical Resource. Design and Implementation of a Graph-based Lexicon Model*. Walter de Gruyter, Berlin & Boston, MA.