

# Chapter 17

## Word Formation Rules and the Generative Lexicon: Representing Noun-to-Verb Versus Verb-to-Noun Conversion in French

Fiammetta Namer and Evelyne Jacquey

### 17.1 Introduction

The issues we address in this paper focus on the interface between lexical semantics and morphology. Just as lexical semantics can be viewed from a compositional point of view, morphology (that is, morpho-semantics) can be understood as a compositional semantics-constrained mechanism according to Corbin's approach (Corbin 1987, 2001). These two levels of description give us distinct types of intrinsic information on the semantic content of derived words. One way to establish links between the two levels of description is to choose a common formalism for their representation.

Basing our work on the study of the prefixation by M. Aurnague and M. Plénat (1996, 1997), limited to the popular prefix *é-* in French, we (Jacquey and Namer 2003; Namer and Jacquey 2003) have suggested modelising the semantic role of this prefix within the framework of the Generative Lexicon Theory (Pustejovsky 1995). In this paper, we further this approach to modelisation of word formation mechanisms and apply this modelisation to account for another word formation (WF) process type in French, namely the NtoV versus VtoN conversion. Our reasons for focusing on the phenomenon of conversion are threefold.

First, it consists in a non-conventional mechanism because of the absence of any affix. The absence of such morphological mark is crucial as far as conversion orientation is concerned. On the basis of the pairs *dance<sub>N</sub>/dance<sub>V</sub>* and *butter<sub>N</sub>/butter<sub>V</sub>* for example, what has to be decided is whether the (output) verb is converted from the (input) noun or vice-versa. This decision will allow us to draw the definition of the output with respect to that of the input.

---

F. Namer (✉) • E. Jacquey  
UMR 7118 ATILF CNRS & Université de Lorraine, Nancy, France

Second, verb and noun pairs related by the conversion process is of interest not only for WF research, but also for the lexical semantics. For instance, Government and Binding-oriented literature sometimes refers to it as to *noun incorporation* phenomenon (Hale and Keyser 1993), while Generative Lexicon Theory proposes to characterise their structure with the so-called *shadow argument* (Pustejovsky 1996).

Finally, this WF process is a multilingual phenomenon. It is both productive and frequent, and found at least in French and English.

Focusing on the French language, our formalisation proposal aims to account for the following aspects, as it will be discussed in this chapter: (1) a corpus-based analysis producing 2,500 homograph Noun/Verb pairs; (2) a ranking of these pairs according to semantic criteria; (3) a modelisation proposal stemming from the analysis of the most frequent and productive classes.

The rest of the chapter is structured as follows: Sect. 17.2 summarises briefly why and how we propose to modelise WF processes within the Generative Lexicon Theory; Sects. 17.3 and 17.4 focus on the conversion process itself. Section 17.3 compares hypotheses coming from linguistic theoretical studies with empirical results obtained by means of the corpus-based analysis mentioned above. Section 17.3 ends by a synthetic table which ranks conversion classes as a result of this comparison. According to these classes, Sect. 17.4 finally suggests two formal models for NtoV and VtoN conversion processes, respectively.

## 17.2 Word Formation Modelisation Within GL: MS-CS

This section focuses on modelisation backgrounds. First, Sect. 17.2.1 gives the linguistic theoretical background WF on which processes rely. Section 17.2.2 deals with the motivation for modelisation itself.

### 17.2.1 Theoretical Background: Corbin's Approach to Word Formation

Among WF theories, research initiated for French in Corbin (1987) provides descriptions that put semantics in the forefront. More precisely, her WF theory is based on three statements:

1. Morphology is autonomous. In other words, the lexicon of the morphologically constructed words is generated by domain specific rules: WF rules and their outputs are independent of e.g. syntactic information. This statement agrees with e.g. (Aronoff 1976);
2. Morphology is regular, i.e. the morphologically constructed words lexicon is regular. Surface exceptions can always be given some explanation whether semantic, phonetic, diachronic, etc.;

3. WF rules associate several kinds of constraints, phonetic, semantic and categorial ones being the most important. It has been established that categorial conditions for WF can be derived from semantics (Corbin 2001; Dal 1997).<sup>1</sup>

Consequently, Corbin's theory foresees that part of the lexical meaning of a morphologically constructed word, called "lexical constructed meaning", is built together with its constructed form.

Considering WF rules from this theoretical point of view clarifies the relationship between lexical semantics and morphology which rules WF, since a morphologically constructed word is above all a matter of semantic constraints. Constraints are exerted both on the **base** (called here **input**) and on the **derived word** (called here the **output**), through WF processes, which can be suffixation, prefixation, conversion or compounding processes. The lexical meaning of both the input and output are opposed to the meaning of the WF process itself, which can be seen as a computational (or instructional) device. In contrast with an input or an output, a WF process does not "mean" anything, but provides a guideline for the output meaning.

As stated in Sect. 17.1, in order to enable WF rules to be displayed as lexical semantics constraints, one way to proceed is to choose a common formalism of representation. The chosen formalism must be able to express semantic constraints at distinct levels, especially at syntactic and semantic levels, for any WF process. In addition, a given WF process may select only specific aspects of its input meaning, in order to build the meaning of the corresponding output.

The expressivity of the Generative Lexicon Theory (GL) makes it suitable to represent the just mentioned constraints. More precisely, GL is modular enough to integrate a level of morphological description and it is rich enough to constrain both input and output of a given WF process. The next section summarises how the GL-based WF mechanism has been set.

## 17.2.2 Formal Background: Our Approach to WF Modelisation

To achieve the goal of modeling WF in French, two basic approaches can be considered: (a) encoding the affixes themselves or (b) setting up abstract parametrised lexical units describing the outputs. One argument for the first choice would be the fact that affixes can be considered as some type of predicates operating on and controlling both the input and the output, from structural, categorial and semantic points of view. The first approach though is inadequate for two main reasons:

1. Encoding affixes to model WF would mean reducing WF to affix-based processes, and would consequently exclude both compounding and conversion. Keep in

---

<sup>1</sup>Phonetics blocks semantics if one is competing with the other. This phenomenon will not be addressed in this paper, since we focus only on WF semantic constraints: for a detailed description, see Plénat and Roché (2004).

mind that the latter consists, loosely speaking, in building new words by means of a simple part of speech (POS) change<sup>2</sup>;

2. The very nature of affixes is another counterargument. According to the morphological theory our study is based on, an affix does not belong to any of the major POS categories. In addition, it bears no referential meaning: consequently, it does not seem logical to modelise its semantic content since it has no proper semantic content.

Thus, in previous studies (Jacquy and Namer 2003; Namer and Jacquy 2003), we turned to the second approach: namely, designing an abstract model which is intended to define the common properties shared by the **outputs** of a given WF process,<sup>3</sup> whatever the involved morphological process. This **abstract lexical unit** (ALU) is instantiated through the **input** content of the WF process, which provides thus the abstract output with distinctive, specific properties.

In order to constrain the combination of an ALU with and only with licensed inputs, we have decided to add a new attribute-value pair at the most ALU embedding level: this pair, encoding the required semantic features of the WF inputs, is referred to as the **morphological structure** (MS).

Finally, in order to instantiate a well-formed constructed word content from the ALU, we assumed one unification mechanism: the **morphological structure composition schema** (MS-CS).<sup>4</sup> Through MS-CS, only the candidate inputs with the appropriate features matching the MS content of the ALU are selected for the formation of well-formed outputs. This unification procedure also entails the instantiation of the right features on the output.

Based on the unification principle, the morphological structure composition schema (MS-CS) in (Fig. 17.1) governs the composition between a given abstract lexical unit and an actual input, in order to build the meaning of a well-formed output. In our conception, MS-CS is meant to be WF process-independent: among its arguments, ALUs are thus likely to represent any WF process, and actual words (both input and output) can belong to any major POS categories.

The MS-CS behavior is twofold. First, when the ALU morphological structure (MS) unifies with the actual input, the success of this unification, noted by the ms index,<sup>5</sup> means that this input satisfies the constraints required by the ALU MS. As a first consequence, relevant features are propagated into the appropriate ALU structures, namely the argumental structure as, the event structure es and the qualia structure qs. Second, the MS-CS schema ensures the propagation of the updated features from the ALU to the output in order to provide the latter with a well-formed semantic content.

<sup>2</sup>See Sect. 17.3 for a more accurate definition of conversion.

<sup>3</sup>i.e. an underspecified referential lexical unit.

<sup>4</sup>For a detailed description of the model, see Jacquy and Namer (2003) and Namer and Jacquy (2003).

<sup>5</sup>As is well-known in the framework of typed feature structures, indices are used to ensure value unification (Shieber 1986).

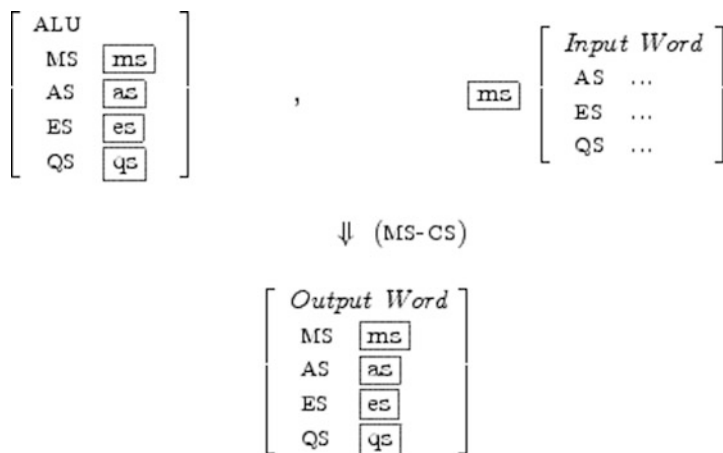


Fig. 17.1 Composition Schema (MS-CS)

The role played by our morphological device, in which MS-CS interact with an abstract lexical unit, can be viewed as a lexical semantics-driven modelisation of the so-called ‘word-based’ model in morphological theory. The choice of such a model (Bybee 1988; Koenig 1999; Fradin 2003) is opposed to ‘morpheme-based’ model (Selkirk 1982; Di Sciullo and Williams 1987; Lieber 1992). Unlike the latter, the word-based model is suitable for the description of non-concatenative word formation processes. In word-based models, the relationship between complex words is captured by formulating **word-schemas** which represent the common features of sets for morphologically related words. According to (Haspelmath 2002:47), “a word-schema is like a lexical entry in that it contains information on pronunciation, syntactic properties and meaning, but it may contain variables. In this way, it abstracts away from the differences between the related words and just expresses the common features”. A word-schema **subsumes** a set of words, that in turn **match** given schemas. Morphological relationships are therefore represented by correspondences between word-schemas. Word pairs that match correspondent schemas are thus related by a particular morphological relationship.

Hence it can be easily seen that ALUs formalise word-schemas and MS-CS represents morphological correspondences between schemas, which inputs and outputs have to match in order to instantiate actual morphologically related word pairs.

### 17.3 Data and Linguistic Description

Our aim is to reuse the MS-CS approach just discussed, in order, this time, to formalise the so-called conversion morphological process in French. This section covers the arguments for choosing this particular word formation type, together with

a brief summary of its main linguistic theoretical properties. Next we will see how a corpus-based experiment is used to match these theoretical properties against large-scale observed characteristics. Finally a set of the most frequent, productive, and stable linguistic properties of noun to verb and verb to noun conversions results from this comparison.

### 17.3.1 *Issues with Conversion*

The morphological conversion process produces an output lexical unit (the convert) from an input lexical unit belonging to a different syntactic category (the base), without any morphological mark. The only visible mark on the output belongs to the inflectional paradigm characterizing its category.<sup>6</sup> In French, verbs (V) may be converted from nouns (N) (*balai*<sub>N</sub> [*broom*<sub>N</sub>] →<sub>conv</sub> *balay(er)*<sub>V</sub><sup>7</sup> [*sweep*<sub>V</sub>]), or from adjectives (A) (*vide*<sub>A</sub> [*empty*<sub>A</sub>] →<sub>conv</sub> *vid(er)*<sub>V</sub> [*empty*<sub>V</sub>]); N may be converted from verbs (*vol(er)*<sub>V</sub> [*fly*<sub>V</sub>] →<sub>conv</sub> *vol*<sub>N</sub> [*flight*<sub>N</sub>]) or from adjectives (*portable*<sub>A</sub> [*portable*<sub>A</sub>] →<sub>conv</sub> *portable*<sub>N</sub> [*laptop*<sub>N</sub>]), the opposite rarely being true (*orange*<sub>N</sub> →<sub>conv</sub> *orange*<sub>A</sub>). Being unmarked,<sup>8</sup> this type of word formation entails the issue of the process orientation<sup>9</sup> i.e. there is no formal way to decide which one of N or V is the conversion output in e.g. *balai*, *balay(er)*, *vol* and *vol(er)*. Within the chosen WF theoretical approach, answering the orientation question amounts into making semantically driven decisions. In other words, detecting e.g. the **NtoV** versus **VtoN** conversion orientation means classifying Noun/Verb (quasi)homograph pairs according to a semantic relation.

### 17.3.2 *NtoV Versus VtoN*

The choice of focusing on the Noun/Verb pairs has been motivated by the presence of a large amount of such pairs, and by the high interest they gather within the

<sup>6</sup>Some authors, especially those who belong to the Government and Binding tradition, call conversion by means of the term *zero affixation*.

<sup>7</sup>We choose to represent verbs by means of their infinitive form: the inflection ending marking the V infinitive *-er* is put between brackets. Some phonological variations, such as here  $\langle \varepsilon \rangle \rightarrow \langle \varepsilon j \rangle$ , graphically realized by the [ai] → [aj] alternation, may occur within the conjugation paradigm.

<sup>8</sup>Inflectional endings, such as infinitive *-er*, are surface marks which have nothing to do with word formation.

<sup>9</sup>When conversion involves morphologically complex lexical units, orientation may be decided from the input and output internal structure. For instance, the suffix *-ure*, building deverbal nouns, appears in the pair *hachure*<sub>N</sub>/*hachur(er)*<sub>V</sub> [*hatching*, *zebra marking*<sub>N</sub>/*hatch*<sub>V</sub>], forcing the N →<sub>conv</sub> V orientation. In fact, the other orientation would imply the output noun be formed both by conversion, from *hachur(er)*, and by suffixation, from *hach(er)* [*chop*, *cut*, *mince*<sub>V</sub>], which is not possible since *hachure* is not polysemous. See Namer (2003), which follows Corbin (1987), for a conversion typology according to input/output morphological complexity.

linguistic community. In fact, according to Corbin (2004), there are no cases of  $V \rightarrow_{\text{conv}} A$  in French. Moreover the  $N \rightarrow_{\text{conv}} A$  type is limited to the production of chromatic adjectives derived from nouns referring to fruits or flowers (*rose*, *orange*...), and to the production of behavior adjectives converted from nouns referring to stereotypical animals (*bête* [*beast<sub>N</sub>*], *cochon* [*pig<sub>N</sub>*] ...). Conversely, both  $N \rightarrow_{\text{conv}} V$  and  $V \rightarrow_{\text{conv}} N$  have been observed, in French as in other languages, even though morphology researchers (at least, the authors whose results are briefly reported below) do not often agree as far as conversion orientation is concerned.

A second argument, directly related to the first one justifying our choice is the semantic heterogeneity of verbs and nouns involved in conversion processes. Regardless, for the time being, of their possible role of input or output in the conversion process, let us notice that nouns may denote concrete (*sucre<sub>N</sub>/sucr(er)<sub>V</sub>* [*sugar<sub>N/V</sub>*]), animate (*singe<sub>N</sub>/sing(er)<sub>V</sub>* [*monkey<sub>N</sub>/mimic<sub>V</sub>*]), human (*guide<sub>N</sub>/guid(er)<sub>V</sub>* [*guide<sub>N/V</sub>*]), or abstract entities (*nage<sub>N</sub>/nag(er)<sub>V</sub>* [*swimming<sub>N</sub>/swim<sub>V</sub>*]); that verbs may describe instrumental (*hache<sub>N</sub>/hach(er)<sub>V</sub>* [*axe<sub>N</sub>/chop<sub>V</sub>*]), dissociative (*plume<sub>N</sub>/plum(er)<sub>V</sub>* [*feather<sub>N</sub>/pluck<sub>V</sub>*]), or locative (*coffre<sub>N</sub>/coffr(er)<sub>V</sub>* [*chest<sub>N</sub>/throw inside<sub>V</sub>*]) processes; and that they may belong to all kinds of eventualities: activities (*crayon<sub>N</sub>/crayonn(er)<sub>V</sub>* [*pencil<sub>N</sub>/scribble<sub>V</sub>*]), transitions (*transport<sub>N</sub>/transport(er)<sub>V</sub>* [*transport<sub>N</sub>/carry<sub>V</sub>*]), etc.

Last but not least, our interest in the Noun/Verb pairs is related to the fact that their linguistic description bridges together word formation and lexical semantics. Given that the Noun/Verb orientation is exclusively a matter of semantics, deciding for  $N \rightarrow_{\text{conv}} V$  or for  $V \rightarrow_{\text{conv}} N$  amounts to detecting the semantic properties on  $V$  and/or on  $N$ . This is in order (1) to check which of  $V$  or  $N$  is obtained from the other one, and, consequently, (2) to determine the semantic relationship holding  $N$  and  $V$ . This second point amounts to draw the definition of the output word by means of the input meaning. From these results, a (first attempt of a) semantic-based typology of  $N \rightarrow V$  and  $V \rightarrow N$  conversion should emerge, as we shall see below.

### 17.3.3 Theoretical Assumptions

Apart from the attempt of orienting  $N \rightarrow V$  from  $V \rightarrow N$  conversion according to **phonological marks** (see e.g. Katamba 1993<sup>10</sup>), literature regarding Noun/Verb conversion tries to give semantic motivations to their classification proposals. For Aronoff and alii (1984), the orientation has to do with **thematic roles attached to  $V$** , and to which role  $N$  may, or may not play. For Mel'cuk (1996, 1997), some  $V \rightarrow N$  conversions are what he calls empty categorial conversions, which may occur between an input lexical unit and an output lexical unit with **stronger distributional constraints** than those of the input. On the other hand, non-empty categorial

<sup>10</sup>His approach, that follows e.g. Marchand (1969), Adams (1973), Kiparsky (1982) is not applicable for French. Also it is not useful for providing outputs with a definition involving their input.

conversions are generally oriented according to the **semantic inclusion relation** between the involved lexical units X and Y: if the meaning of X is included in that of Y, then  $X \rightarrow Y$ . Moreover, he proposes, following (Corbin 1987), a **paradigmatic orientation**<sup>11</sup> of Noun/Verb conversion: it is oriented in the same way as affixation with the same semantic relation. For instance, since *-eur* in French basically builds agents ( $nag(er)_V \rightarrow_{-eur} nageur_N$  [ $swim_V/swimmer_N$ ]), and no other affix involves agents, all Noun/Verb pairs exhibiting an “agent” semantic relation should belong to that paradigm, and thus, for instance  $guid(er)_V \rightarrow_{conv} guide_N$ .

Among the assumptions briefly reported above, the **paradigmatic orientation** hypothesis seems to be the most promising: in fact, Aronoff’s relying on **thematic roles** would require a clear, stable and homogeneous definition of them, which is unfortunately not the case. As for Mel’cuk, he is neither able to define formally **distributional constraints** (which rules empty VtoN categorial constraint) nor **semantic inclusion** (which rules non-empty NtoV conversion).

However, paradigmatic orientation hypothesis is not a completely satisfactory solution. First, it does not account for pairs such as  $babouin_N/babouiner_V$  [ $baboon_N/act\ as\ a\ baboon_V$ ], in which **imitation verbs** depict the referent of the agent as acting in the same way as the referent of the base noun they are morphologically constructed from. Second, it leads to contradictory situations, e.g. when nouns refer to instruments. According to the paradigm, the conversion relation of Noun/Verb pairs should be  $V \rightarrow_{conv} N$  oriented when N denotes an instrument, since the only affixation process dealing with instruments in French are suffixes *-oir* and *-eur*, which both form deverbal nouns. Therefore, for *-oir*, we have for instance  $hach(er)_V \rightarrow_{-oir} hachoir_N$  [ $chop_V/chopper_N$ ]. But for the same input, we notice that we also have  $hach(er)_V \rightarrow_{conv} hache_N$ , [ $axe_N$ ] bearing (apparently) the same semantic relation. This is also the case with other N/N-*oir* or N/N-*eur* pairs:  $drain_N/draineur_N$  [ $drain_N/drainer_N$ ],  $gril_N/grilloir_N$  [ $gril_N/griller_N$ ]. The meaning variation between the compared nouns may indicate that Noun/Verb and N-*oir*/V or N-*eur*/V do not belong exactly to the same paradigm. A clear example of this is the case of the the verb  $agraf(er)_V$  [ $staple_V$ ]: the noun  $agrafe_N$  [ $staple_N$ ] refers to the concrete entity that performs the process itself; and the noun  $agrafeuse_N$ <sup>12</sup> [ $stapler_N$ ], the instrument which must be used so that these staples can do their job. If the instrumental paradigm cannot always be clearly stated, then there is no longer much evidence for the  $V \rightarrow_{conv} N$  orientation, when N is an instrument. Furthermore, D. Corbin partially reconsiders in later papers the overall paradigmatic hypothesis (Corbin 1997, 2004), mentioning instruments and instrumental verbs ( $scie_N/sci(er)_V$  [ $saw_N/V$ ]) as NtoV conversion cases.

<sup>11</sup>also called ‘overt analogue’ principle (Sanders 1988).

<sup>12</sup>Actually, *-eur/-euse* are nothing but two gender variations of the same affix: *-euse* is a possible feminine form corresponding to *-eur*. We assume that gender variation has to do with inflection, and thus is not a matter of WF, at least for nouns referring to inanimate entities, see (Corbett 1991) for an accurate discussion on this issue.



Be that as it may, we shall keep this paradigmatic assumption as a starting point. In addition to the NtoV WF processes, this hypothesis has also been the theoretical foundation for VtoN descriptions and analyses. One of the main contributors to these studies for French is F. Kerleroux: a very detailed analysis of converted deverbal nouns' properties has been carried out by Kerleroux (1996a, b, 1997, 1999). Furthermore, Kerleroux (2004), Fradin and Kerleroux (2003a, b) redefine the notion of a lexeme. Consequently, they draw a set of conditions constraining VtoN conversion, using the differences these authors record between conversion and apocope, from both distributional and semantic points of view. VtoN conversion is also the object of study in Meinschäfer (2003); J. Meinschäfer proposes a set of criteria predicting the deverbal noun argument structure, with respect to that of the input verb. More precisely, she shows that deverbal nouns share with their input verb their aspectual and argumental properties, provided that the verb is not a causative event: *Max recule la chaise* [*Max moves back the chair*] → *\*Le recul de la chaise (par Max)* [*(Max's) moving back of the chair*]. However, she observes that causative and non-causative readings always alternate, thus allowing conversion: *Max recule* [*Max is going back*] → *le recul de Max* [*Max's backward movement*].

To sum up Noun/Verb conversion orientation possibilities, we can make the following assumptions:

- $N \rightarrow_{\text{conv}} V$  holds when (a) N is itself morphologically constructed (*hachure* [*hatching<sub>N</sub>*]), (b) N is an instrument/substance used to perform the process described by V (*scie* [*saw<sub>N</sub>*], *sucré* [*sugar<sub>N</sub>*]), (c) N is the place where the process is performed (*coffre* [*chest<sub>N</sub>*]), (d) N is the stereotypical agent of the process (*singe* [*monkey<sub>N</sub>*]);
- $V \rightarrow_{\text{conv}} N$  holds, basically, when N is abstract; so N describes the process, its result or its product (*transport*); besides, N may denote the process agent (*guide*).

With this first classification in mind, let us now turn to the corpus analysis. We have collected a set of Noun/Verb homographs pairs, in order to (a) try to classify them according to the above criteria, (b) if (a) is not possible, to define new classes.

### 17.3.4 Corpus

To check the validity of the linguistic hypotheses performed above, we have collected the set of quasi-homograph verbal and nominal lexical units from a large-scale machine readable dictionary, mainly the TLFnome<sup>13</sup> word list. Lexical units

<sup>13</sup>TLFnome is a lexicon of inflected forms developed at the INaLF based on the nomenclature of the *Trésor de la Langue Française*, a general language multi-volumes dictionary. It currently contains 63,000 lemmas, 390,000 forms and 500,000 entries. It is in the course of being supplemented by 36,400 additional lemmas from the TLF index.

are labeled with the appropriate part-of-speech, and have at worst different endings, and allomorphic variants (e.g., changing thematic vowel aperture, graphically marked by a diacritic, e.g. with /B/→/ε/in *relev(er)<sub>V</sub>/relève<sub>N</sub>* [*pick up<sub>V</sub>/relief<sub>N</sub>*], or by doubling consonant e.g. with /T/→/Cn/in *collision<sub>N</sub>/collisionn(er)<sub>V</sub>* [*collision<sub>N</sub>/collide<sub>V</sub>*]).<sup>14</sup> A set of 2,500 Noun/Verb pairs have thus been gathered, half of which have been manually verified. The verification objectives are the following:

- checking whether the paired elements are actually linked by conversion;
- deciding for the conversion orientation, according to: (1) the theoretical assumptions given in Sects. 17.3.3 and 17.2 definitions and etymologies provided within dictionaries;
- if needed, proposing new classes, or constraining the existing ones.

The conclusions of this large-scale verification are summed up in Sect. 17.3.5, from both a qualitative (ranking Noun/Verb pairs with respect to semantic classes) and quantitative (classifying Noun/Verb pairs according to their frequency) point of view.

### 17.3.5 Synthesis

Tables 17.1 and 17.2 below summarize the observations resulting from corpus data analysis. First, as far as **VtoN** conversion conditions are concerned (Table 17.2), the results are all in all in conformity with the hypotheses made in the previous section. Nouns massively refer to abstract entities (class -2-), although semantic derivations are sometimes observed: for instance, the process noun *applique* (*[application<sub>N</sub>]*) converted from *appliquer* (*[apply<sub>V</sub>]*) has a specialised meaning which leads this noun to refer to (concrete) entities, “whose function is to be fixed/mounted/hung (onto the wall)” namely *wall lamps*.

Concerning **NtoV** (Table 17.1), there are discrepancies between theoretical assumptions and corpus analysis results, which makes the definition of new classes. For instance, similarity verbs are not only met with respect to the agent (*singer<sub>V</sub>*), but also with respect to the theme, which is affected by the change-of-state transition process described by the verb: *marbr(er)<sub>V</sub>* [*marble<sub>V</sub>*] (class -8-). The property acquired by the theme is a shape, a color, etc. described by the referent of the input noun. In addition to class -2-, grouping artefactual instruments/substance-based verbs, another set of rather similar verbs, has been collected in the so-called class -2’-: the input noun, referring to a part of the body

---

<sup>14</sup>Due to the technique we used to collect our corpus, N/V pairs exhibiting strong allomorphy or suppletive variations are not included in our study (e.g. *pleuv-oir<sub>V</sub>/pluie<sub>N</sub>* [*rain<sub>V/N</sub>*]). A manual checking should be necessary to confirm their behaviour to be in conformity with the obtained conclusions (Sect. 17.3.5, Table 17.1).

**Table 17.1** NtoV conversion classes

Class	N → V	Nb	% <sup>a</sup>
		<b>789</b>	<b>73.6</b>
1	N is a morphologically constructed lexical unit: <i>hach/hachurer</i> [ <i>hatching/hatch</i> ] <i>N</i> = “action of <i>V</i> ”	86	<b>8.02</b>
2	N refers to either a typical instrument ( <i>drain/drain(er)</i> ) or to the substance ( <i>sucré/sucr(er)</i> ) implied as a mean to realise the process described by V: <i>V</i> ( <i>N<sub>theme</sub></i> ) = “do smthng to <i>N<sub>theme</sub></i> using <i>N</i> ”	<u>372</u>	<b>34.73</b>
2'	N refers to either a part of human body or a human characteristics ( <i>cil/cill(er)</i> , <i>raison/raisonn(er)</i> ) <i>N<sub>agent</sub></i> <i>V</i> = “Use <i>N</i> , which (is part of/characterizes) <i>N<sub>agent</sub>[+hum]</i> ”	14	<b>1.3</b>
3	N refers to a stereotypical agent of the process: <i>singe/sing(er)</i> <i>V</i> = “(do what <i>N</i> would do behave as <i>N</i> )”	<u>69</u>	<b>6.44</b>
4	N refers to the place or temporal interval in which the process takes place: <i>coffre/coffr(er)</i> <i>V</i> ( <i>N<sub>theme</sub></i> ) = “do or put <i>N<sub>theme</sub></i> (with)in/during <i>N</i> ”	<u>50</u>	<b>4.66</b>
5	N is a meteorological phenomenon: <i>neige<sub>N</sub>/neig(er)<sub>V</sub></i> [ <i>snow<sub>NV</sub></i> ] ( <i>V</i> is impersonal) <i>V</i> = “It does <i>N</i> ”	5	<b>0.46</b>
6	N refers to the product obtained by dissociating via V the entity referred to by the V theme: <i>plume/plum(er)</i> <i>V N<sub>theme</sub></i> = “produce <i>N</i> by dissociating <i>N</i> from <i>N<sub>theme</sub></i> ”	34	<b>3.17</b>
7	N refers to a sound, a noise, a (speech) act ( <i>belote/belot(er)</i> , <i>laius/laiuss(er)</i> , <i>peste/pester</i> ) N and V may be borrowings ( <i>crash(er)/crash</i> ) or onomatopoeias lexical units ( <i>blablat(er)/blabla</i> ) <i>V</i> = “do/say/have <i>N</i> ”	<u>92</u>	<b>8.59</b>
8	Either N, or some of its metonymic derived entities ( <i>N<sub>mtny</sub></i> ) is what <i>N<sub>theme</sub></i> looks like at the end of the process described by V ( <i>marbr(er)</i> ) <i>V</i> ( <i>N<sub>theme</sub></i> ) = “provide <i>N<sub>theme</sub></i> with either <i>N</i> or <i>N<sub>mtny</sub></i> characteristics or appearance”	67	<b>6.25</b>

<sup>a</sup>Among the 1,250 collected Noun/Verb pairs, 1,071 are truly related through conversion. The 179 other ones often correspond to homonymy cases, such as *griffon<sub>N</sub>/griffonner<sub>V</sub>* [*scribble<sub>V</sub>*]. Total and percentages are calculated on the basis of these actual 1,071 conversion pairs

**Table 17.2** VtoN conversion classes

Class	V → N	Nb	%
		<b>282</b>	<b>26.4<sup>14</sup></b>
1	N refers to the verb agent: <i>guid(er)/guide</i> <i>N</i> = “the one who <i>V</i> ”	10	<b>0.93</b>
2	N refers to the process described by V, or its result, or its product: <i>vol(er)/vol</i> , <i>recul(er)/recul</i> <i>N</i> = “action   result of <i>V</i> ”	<u>272</u>	<b>25.39</b>

(*cil*<sub>N</sub> [*eyelash*<sub>N</sub>]) or a human characteristic (*raison*<sub>N</sub> [*reason*<sub>N</sub>]) is namely used as an instrument (*cill(er)*<sub>V</sub> [*blink*<sub>V</sub>], *raisonn(er)*<sub>V</sub> [*reason*<sub>V</sub>]). Finally, a ‘default’, heterogeneous class has been drawn, grouping together Noun/Verb pairs in which N may refer to speech acts (*laiüs*<sub>N</sub> vs *laiüss(er)*<sub>V</sub> [*long winded speech*<sub>N</sub>/*expatiate*<sub>V</sub>]), to noise or sounds (*vacarme*<sub>N</sub> vs *vacarmer*<sub>V</sub> [*uproar*<sub>N</sub>/*make an uproar*<sub>V</sub>], *clic*<sub>N</sub> vs *cliquer*<sub>V</sub> [*click*<sub>N/V</sub>]), to concrete action results (*sieste*<sub>N</sub> vs *siester*<sub>V</sub> [*nap*<sub>N</sub>/*have a nap*<sub>V</sub>], *balafre*<sub>N</sub> vs *balafre*<sub>V</sub> [*cut, slash, gash*<sub>N/V</sub>]). This class, labeled with -7-, is defined by means of a shallow link: V = “do/say/have N”. Within this class are also listed N/V pairs where V denotes delocutive acts: *choucou*<sub>N</sub> vs *chouchouter*<sub>V</sub> ([*darling*<sub>N</sub>/*pet*<sub>V</sub>]), *peste*<sub>N</sub> vs *pester*<sub>V</sub> ([*heavens!*<sub>N</sub>/*curse*<sub>V</sub>]). The ‘delocutive derivation’, originally introduced in Benveniste (1966) has been investigated in Cornulier (1976) and Anscombe (1979). Delocutive denominal verbs can be glossed by “To say « N »”. Recently, an historical review of this notion has been described in Larcher (2003).

Moreover, a productive class has been isolated, namely that of borrowings (*crash(er)*<sub>V</sub>/*crash*<sub>N</sub>) and onomatopoeias (*blablat(er)*<sub>V</sub>/*blabla*<sub>N</sub> [*waffle on*<sub>V</sub>/*waffle*]). As nouns belonging to these N/V pairs denote concrete entities (sounds and (speech) acts), they have been included in class -7-.

In addition to both the initial linguistic assumptions and the newly discovered classes, Tables 17.1 and 17.2 also includes both new columns with quantitative results obtained from the dictionaries corpus analysis, and new cells, corresponding to the new discovered semantic classes just described.

Whereas VtoN conversion appears to be a stable WF process, leading to the formation of almost only abstract nouns, characterising NtoV types is a much less straightforward task. In fact, for this purpose, we have examined input N (formal, semantic, etymologic) features only. To refine this classification, a next step will be to compare these criteria with output verbs properties.

According to these (though perfectible) results, we can model the most frequent and seemingly productive conversion classes. With this choice, classes -2’-, -5- and -6- in Table 17.1, together with class -1- in Table 17.2, are excluded. Furthermore, we have chosen to disregard heterogeneous cases (i.e. classes -7- and -8- in Table 17.1) at the time being, the linguistic content of this set of nouns and verbs having in fact to be further examined; in particular, in Sect. 17.4.1.3, we come back to the reasons why Noun/Verb pairs which are members of class -8- are not accounted for in this chapter. Finally, the last excluded class is class -1-, Table 17.1, since NtoV orientation is in this group purely structure-driven. These decisions amount to design two ALUS, the former constraining and producing denominal converted verbs, the latter defining the basic structure of deverbal converted nouns. In Sect. 17.4, we shall see which of the input properties can be encoded within ALUS, which ones fall within the competence of the actual input, and how the MS-CS mechanism is able to build the right output representations, whatever the requested Noun/Verb class.

## 17.4 Modeling

As announced in Sect. 17.2, the formal representation we wish to obtain combines the following requirements: (1) MS-CS is taken as an input to output unification mechanism, (2), a unique **NtoV** unified  $[X_N]_{V}ALU$  records the linguistic constraints common to classes -2-, -3- and -4- in Table 17.1, while a unique **VtoN**/ $[X_V]_{N}ALU$  does the same for the representation of class -2- in Table 17.2 (see Sect. 17.3.5). Behind the idea of accounting for regular, productive and frequently represented conversion classes, the goal is to predict the characteristics of the most likely Noun/Verb conversion producted neologisms.

### 17.4.1 Noun-to-Verb

Examining Table 17.1, Sect. 17.3.5, and excluding class -8-, three **NtoV** classes are very productive: class -2- ( $V = \text{“do something using } N\text{”}$ ), class -3- ( $V = \text{“}(do\ what\ N\ would\ do|behave\ as\ N)\text{”}$ ) and class -4- ( $V = \text{“do or put something (with)in/during } N\text{”}$ ). As we shall see in the Sect. 17.4.1.1, all output verbs are based on a unique ALU called  $[X_N]_V$ . Section 17.4.1.2 focuses on some examples for each of the classes which has been taken into account.

#### 17.4.1.1 $[X_N]_V$ Abstract Lexical Unit for Noun-to-Verb Conversion

The following ALU in (Fig. 17.2) accounts for the way output verbs inherit properties from the appropriate input nouns:

- They inherit relevant argumental properties from their input noun, namely only those parameters which are used in input noun qualia roles  $\boxed{2}$  and which are inherited by the verb. These parameters are encoded by  $\boxed{a_i}$  variables;
- They inherit relevant aspectual and event structure parameters from their input noun, namely only those parameters which are used in the input noun qualia roles  $\boxed{2}$  and which are inherited by the verb. These parameters are encoded by  $\boxed{e_j}$  variables;
- They inherit only a part of the semantic content of their input noun, represented here by a part of the noun qualia. Mutual disjunctions ( $\oplus$ ) rule out overlapping between classes which have been accounted for:
  - **if** the input noun denotes an artefact (class -2-) or a location (class -4-), **then** the qualia of the output verb consists only in the telic value of the input qualia  $\boxed{4}$ ,
  - **if** the input noun denotes a natural entity (class -3-), **then** the output verb inherits only the agentive value in the formal quale  $QS|FORM|AG$  of the input

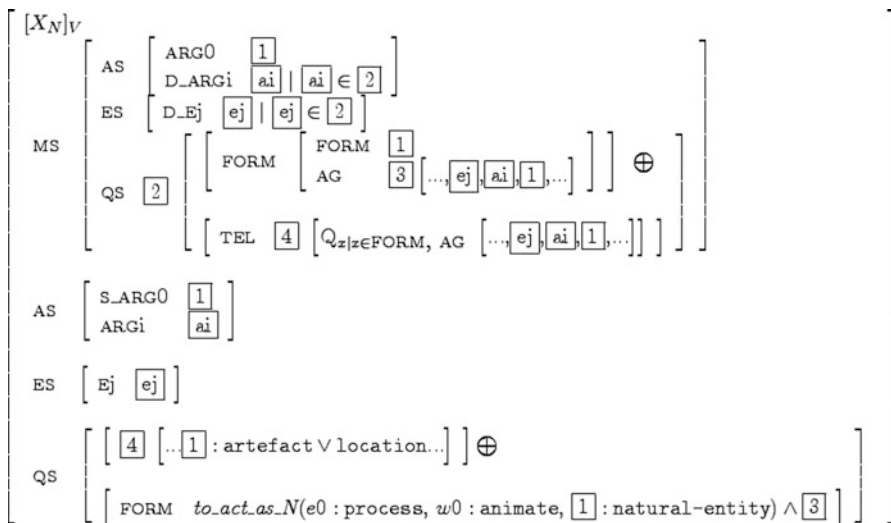


Fig. 17.2  $[X_N]_V$  ALU

[3], and as a consequence, the qualia of the output verb consists in this case in a formal role whose value is the conjunction of the predicate *to act as N* and the QS|FORM|AG value, if any.

### 17.4.1.2 Some Examples

As we shall see with the examples below,  $[X_N]_V$  combined with the appropriate input noun enables the representation, via MS-CS, of each sort of output verbs from the following **NtoV** conversion classes: class -3-, with imitation verbs like *sing(er)*, class -2-, with instrumental verbs like *drain(er)*, *crayonn(er)*, *dynamit(er)* [*dynamite<sub>V</sub>*] and class -4-, with locative verbs like *usin(er)* [*manufacture<sub>V</sub>*], *coffr(er)*. Examples from each class are meant to illustrate various cases of inherited aspectual properties.

#### Imitation Verbs

As said before, class -3- imitation verbs are built from nouns which denote natural entities, e.g. *singe*. Let us see how the output verb *sing(er)* is produced from its nominal input *singe* (Fig. 17.3). First, we may notice that the input noun qualia structure indicates that *singe* is an animal [5] bearing a prototypical behavior i.e. imitating a model [6]. This behavior is propagated through [6] and via  $[X_N]_V$  onto the constructed verb FORMAL role. Therefore, the MS-CS-driven combination

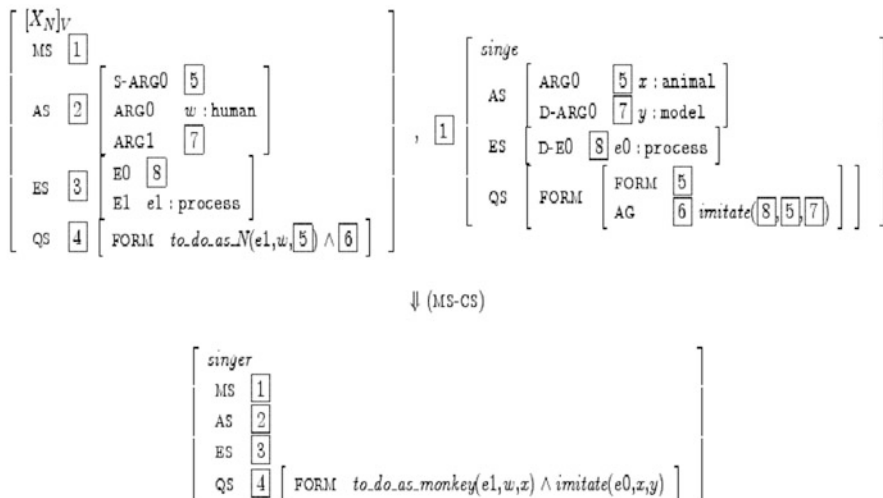


Fig. 17.3 Conversion class -3-:  $\textit{singe}_N \rightarrow_{\text{conv}} \textit{singer}_V$

between  $[X_N]_V$  and the lexical properties of the input noun entails the output verb to be provided with a qualia structure that can be paraphrased by:  $X \textit{sing}(er) Y = "X \textit{acts as a } N\_monkey \textit{ AND monkeys imitate } Y"$ .

### Instrumental Verbs

Prototypically, instrumental verbs are morphologically constructed from input nouns referring to artefacts. Let us consider for instance, the example of  $\textit{drain}_N/\textit{drain}(er)_V$  pair (Fig. 17.4). The output verb  $\textit{drain}(er)$  inherits the TELIC role from its input noun  $\textit{drain}$ , because of the artefactual nature of the noun referent. This TELIC value is a complex structure which is characterised by the qualia label  $\textit{transition\_lcp}$  and which consequently contains the specific features for transitions. Since  $\textit{drain}(er)$  describes an instrumental predicate, its meaning, carried through index  $\boxed{4}$ , can be expressed through the following gloss:  $X \textit{drain}(er) Y = "X \textit{uses } N\_drain \textit{ to extract } Y \textit{ from } Z \textit{ AND } Y \textit{ is extracted from } Z"$ .

In addition, the inheritance of argumental and aspectual properties follows the general principles of noun and verb descriptions in GL. Except for the denoted entity  $\boxed{1}$ , input nouns argumental parameters are always encoded as default arguments (MS|AS|D-ARGi), whereas they are inherited as true arguments in the output verb argumental structure (AS|ARGi). In the same way, the default evenemential parameters in MS|ES|D-Ej are inherited as true parameters in output ES|Ej.

Following the lexical shadowing principle, the argumental parameter  $\boxed{1}$  which encodes the entity denoted by the input noun is displayed as a shadow argument (S-ARG0) in the output verb AS.

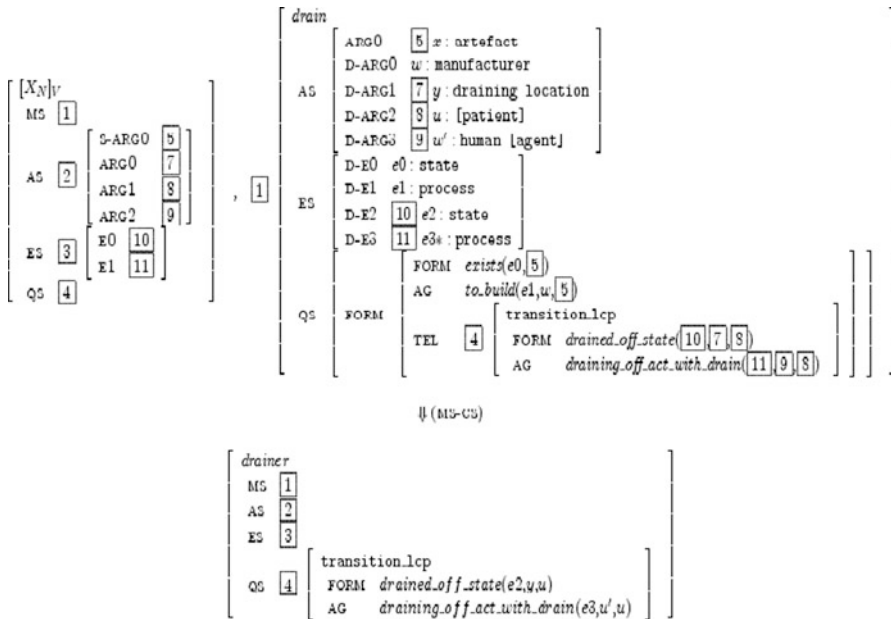


Fig. 17.4 Conversion class -2-:  $drainer_N \rightarrow_{con} drain(er)_V$

The same mechanism is at play with verbs *crayonn(er)* and *dynamit(er)* excepted that here the TELIC role of the input noun *crayon* (resp. *dynamite*) encodes an activity (resp. an achievement). This TELIC value is reflected within the inherited qualia structure of each corresponding output verb: *crayonn(er)* denotes an activity whereas *dynamit(er)* describes an achievement.

### Locative Verbs

As shown in Fig. 17.5, the fact that the input noun *coffre* refers to a place [5] leads to the morphological formation of the locative verb *coffr(er)*. The meaning of *coffr(er)* can be paraphrased by  $X\ coffr(er)\ Y = "X\ locks\ up\ Y\ in\ N\_chest\ AND\ Y\ is\ locked\ up\ in\ N\_chest"$ . This verb inherits the relevant part from the input noun qualia (i.e. [4], its TELIC value), and those appropriate argumental and evenemential parameters which are linked within this inherited qualia part (the state [9] and the process [10]; the agent [8], the patient [7] and the location [5]). The input noun TELIC value being of type *transition\_lcp*, this label is propagated in order to characterize the output verb qualia structure.

The same is at play with the Noun/Verb pair *usine\_N/usin(er)\_V* excepted for the kind of event which is denoted in the TELIC value of the input noun *usine*. This value is of type *activity\_lcp*, and it also characterises the qualia structure of the output verb.



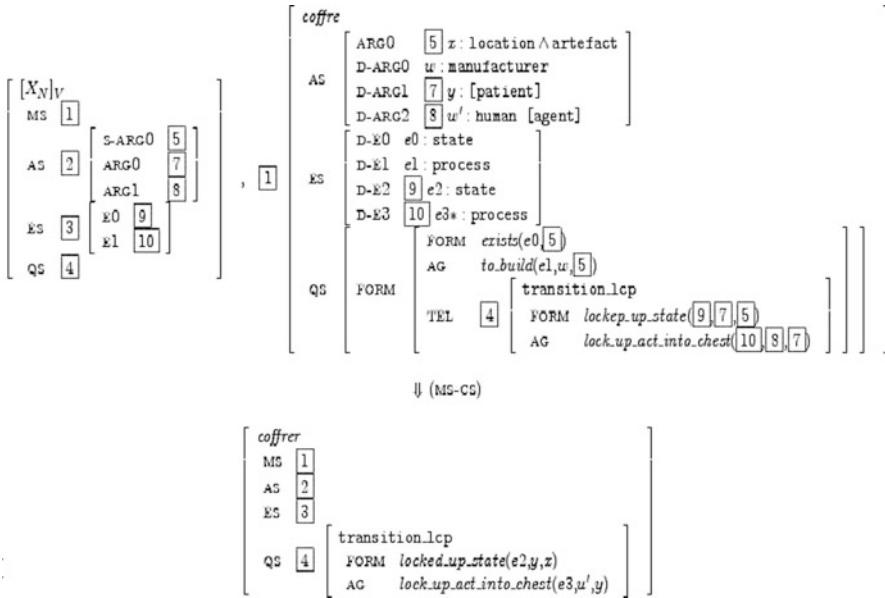


Fig. 17.5 Conversion class -4-:  $coffre_N \rightarrow_{conv} coffr(er)_V$

### 17.4.1.3 Conclusion

This section was devoted to NtoV conversion WF process. We have seen that a unique ALU called  $[X_N]_V$  combined with the appropriate input noun through MS-CS schema is sufficient to build well-formed output verb meanings in a systematic way with respect to the ontological type of input nouns. Three kinds of output verbs are built in this way: imitation verbs like *sing(er)* from input nouns which denote natural entities; instrumental verbs like *drain(er)* from input nouns which denote artefacts; locative verbs like *coffr(er)* from input nouns denoting places or time intervals. As said above, these three kinds of output verbs correspond to three classes of Noun/Verb pairs, respectively class -3- (V = “(do what N would do|behave as N)”), class -2- (V = “do something to  $N_{theme}$  using N”) and class -4- (V = “do or put  $N_{theme}$  (with)in/during N”).

Let us come back to the reasons Table 17.1, class -8- has not been taken into account here. Observing this class, we may notice that a change of state is exerted by the output verb on its theme either with respect to the input noun itself ( $marbre_N \rightarrow_{conv} marbr(er)_V X_{Theme}$ : “ $X_{Theme}$  looks like marble”), or with respect to its shape ( $ballon_N$  [balloon]  $\rightarrow_{conv} ballonn(er)_V X_{Theme}$ : “ $X_{Theme}$  is round as a balloon”), or with respect to one of its parts ( $guêpe_N$  [wasp]  $\rightarrow_{conv} guêp(er)_V X_{Theme}$ : “ $X_{Theme}$  has a wasp waist”), or with respect to its function ( $frégate_N$  [frigate]  $\rightarrow_{conv} frégat(er)_V X_{Theme}$ : “ $X_{Theme}$  is such that its speed is that of a frigate”). In other words, the very meaning within the change of state affecting the referent of the verb

theme, in Noun/Verb pairs belonging to this class, may be a function of one of the input noun qualia roles: e.g. CONST (*guêp(er)*) or TELIC (*frégat(er)*).

Given the evident complexity of these verbs, it seems clear that performing more subtle and discriminating representation of Table 17.1, class -8- verbs would provide us with very interesting results, and therefore deserves further investigation. However, we cannot address this question at the present time because several questions are not answered yet, among which the two of them:

1. No discriminating properties can be exhibited to constrain the membership of a given Noun/Verb pair to the Table 17.1, class -8-, because of the large range of input types: inputs may denote substances (*marbre*), artefacts (*ballon*), animals (*guêpe*), etc.;
2. No discriminating features can be defined to constrain the inheritance of input properties: output verb meaning can be obtained either from that of the whole entity referred to by the input noun, or from that of a related entity: e.g. the shape, some part, the function, etc. of the entity denoted by the input noun.

Answering these crucial issues is a mandatory precursor proposing a formal model for the semantic content of Table 17.1, class -8- output verbs from that of input nouns. A makeshift way to answer the first issue above would be to use some underspecified predicate such as  $V = \text{“to\_give\_some\_characteristics\_of } N\text{”}$ , but such a controversial solution would not solve the second question. As a consequence, we prefer not to account for Noun/Verb pairs of Table 17.1, class -8- as long as points 1 and 2 remain unanswered issues.

## 17.4.2 Verb-to-Noun

As it emerges throughout the section devoted to linguistic descriptions, and according to the quantitative corpus-based values reported on the Table 17.2, Sect. 17.3.5, most deverbal converted nouns (i.e. those labelled by class -2-) describe either the verbal process or its result.<sup>15</sup> A third reading consists in a conceptual or propositional one.<sup>16</sup>

These interpretations are all possible. Some nouns may realise all of them, for instance *marche* [*march<sub>N</sub>/walk(ing)<sub>N</sub>*]: (processive) *la marche durera environ une heure* [*the walk/march will last one hour long*], (result) *la marche des Américaines a été un succès* [*The American women’s march has been a success*], (concept) *la marche est une discipline olympique* [*walking is an olympic sport*]; some other

---

<sup>15</sup>VtoN conversion class -1-, Table 17.2 in which N denotes the agent (*guide, garde*) gathers a non-representative amount of Noun/Verb pairs; therefore we have chosen not to take them into account in our model.

<sup>16</sup>See Pustejovsky (1995:175) for an illustration of this type of denotation with the ambiguous noun *belief*.

nouns have only two interpretations. So *chant* [*song<sub>N</sub>/singing<sub>N</sub>*] is only either resultative *le chant des sirènes a ensorcelé Ulysse* [*The mermaid's song bewitched Ulysses*] or conceptual *le chant est un art* [*singing is an art*]. The MS-CS output lexical unit does not try to guess which of the readings is actually realised by the noun, it just provides nouns with the three possibilities.

#### 17.4.2.1 V and N Minimal Required Features

Gathering the main properties accounted for by various authors (Corbin, Kerleroux, Fradin, Meinschäfer) and mentioned in Sects. 17.3.3, 17.3.4, and 17.3.5, we obtain the following list of minimal requirements the VtoN ALU, noted [*X<sub>V</sub>*]<sub>N</sub>, must satisfy in order to properly constrain the abstract semantic structure of deverbal converted nouns:

1. Its MS – collecting the characteristics all input candidate verbs must share – is as follows:
  - causative readings of input verbs being excluded, the verb qualia label should exclude any potential causative interpretation;
  - the event structure should not be that of a simple state;
  - the argument structure is unconstrained: actually, candidate verbs may or may not be transitive;
2. The description of the noun itself denotes an abstract entity with three possible readings: processive, resultative, or conceptual,
  - in its evenemential (processive or resultative) readings, the output noun inherits all the verbal aspectual and argumental properties, following (Meinschäfer 2003);
  - in its conceptual reading, the output noun refers a priori to a so-called proposition entity. So, this denotation, noted *prop*, must be part of the output qualia label.

Finally, in contrast to what happens for denominal verbs, described in Sect. 17.4.1, and to what Pustejovsky (1996) assumes, input verbs do not carry the output noun index as shadow argument (S\_ARG). We have to remember that Pustejovsky (1996) proposes the S\_ARG value to be instantiated for verbs as *dance<sub>V</sub>* or *butter<sub>V</sub>* by means of what could be considered as an incorporated noun: we agree with this assumption as far as *butter* is concerned: the entity is a logical part of the predicate, but not for *dance*, at least in French. Actually, as for any input verb of a VtoN conversion process, allowing for an S\_ARG value in *dance<sub>AS</sub>* would amount allowing for a circular definition of V and N: the output noun would namely be, at the same time, both morphologically obtained from the verb, and semantically integrated in the verb definition.

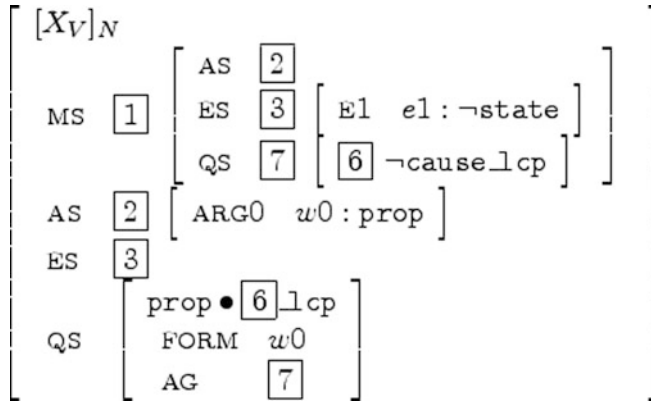


Fig. 17.6  $[X_V]_N$ ALU for VtoN conversion

### 17.4.2.2 $[X_V]_N$ Abstract Lexical Unit

Figure 17.6 below formalizes the set of constraints just recalled in the  $[X_V]_N$ ALU.

For readability sake, input verb AS (resp. ES) is directly coindexed through  $\boxed{2}$  (resp. through  $\boxed{3}$ ) with ALU AS (resp. ES). All the nominal arguments (excepted AS|ARG0 value) are inherited as default arguments, and value sharing would have required a slightly more complex representation.

In the ALU qualia (QS), a new parameter  $w_0$  is used in the FORM value to ensure the existence of a conceptual interpretation of the expected output noun. The MS QUALIA value, characterising the potential input verb, is shared with that of the output noun QS|AGENT'S value through index  $\boxed{7}$ , as soon as this shared value meets the type constraint exerted on the input verb. Recall (Sect. 17.3.3) that, following (Meinschäfer 2003) this type constraint says that V should not have a causative reading, that is identified by the label  $\neg\text{cause\_lcp}$ . The input verb type, represented by the LCP label, and indexed with  $\boxed{6}$ , becomes one component of the output noun qualia label. Given that none of the three potential interpretations of the output noun cooperates in any context, the LCP of this output noun is identified by an exocentric dotted type.<sup>17</sup> This complex type is composed of two simple types: the input verb type indexed with  $\boxed{6}$  and the prop type. Therefore, deverbal converted nouns' LCPs are identified with:  $\text{prop} \bullet \boxed{6} \_lcp$ , where  $\boxed{6}$  stands for any aspectual type, but cause.

<sup>17</sup>Following Pustejovsky (1996), the distinction between endocentric and exocentric dotted types is based on the acceptability of the contextual cooperation between the two readings of a logical polysemic word, these readings being accounted for by two corresponding types which are embedded in the dotted type. The readings of a word associated to an exocentric dotted type do not cooperate, while the two readings of a word associated to an endocentric dotted type do. The exocentric/endocentric distinction seems to be close to that of Cruse (1986) between cooperative versus non cooperative readings of word.

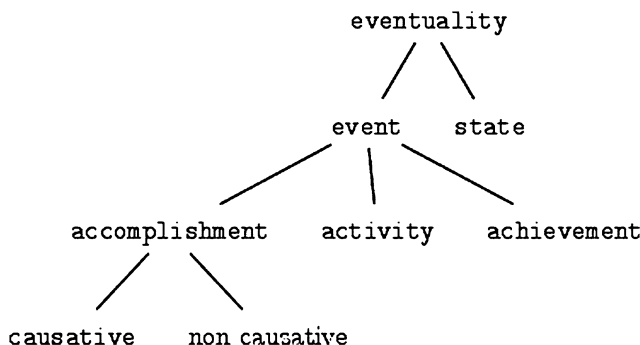


Fig. 17.7 Usual type hierarchy of eventualities

Furthermore, [3] indicates that the type of the first ES event, i.e. E1, cannot be a state. Still according to Meinschäfer (2003), cf. Sect. 17.3.3, this second constraint filters out stative verbs. In other words, deverbal converted nouns denote basically abstract entities, and their AGENTIVE role (in fact, their origin) is the meaning of the verb they are converted from.

As for type accommodation between input verbs and the  $[X_V]_{N\text{ALU}}$  constraints, according to the usually adopted type hierarchy, given in (Fig. 17.7),  $\neg$ cause is equivalent to the entirety of the event subtypes, but cause. Now, cause being an accomplishment subtype, a  $\neg$ cause\_lcp marked verb may express any non-causative accomplishments, activities or achievements.

In the case of a candidate input verb which includes a causal reading and is of an exocentric dotted type, filtering out by unification [6] the verb cause component amounts to keep only its non causative interpretation, by means of those QUALIA role predicates the remaining component type can access. And, consequently, only those evenemential and argument variables used in the accessed predicates are kept in the respective structures.

The whole word formation mechanism, made up with the  $[X_V]_{N\text{ALU}}$ , the candidate input verb **X1**, the MS-CS system and the nominal output noun  $[X_I_V]_N$  is given in (Fig. 17.8).

1. the  $[X_V]_{N\text{ALU}}$  subsumes the common properties of all converted deverbal noun, by defining the minimal requirements on the expected input verb;
2. the potential input verb **X1** has to unify with the ALU MS, in order to activate MS-CS;
3. the actual deverbal noun  $[X_I_V]_N$  results from MS-CS unification process, instantiating  $[X_V]_N$  by means of appropriate **X1** features. Examples of VtoN conversions involving non causative verbs or non causative verb readings, presented in Sect. 17.4.2.3, illustrate this mechanism.

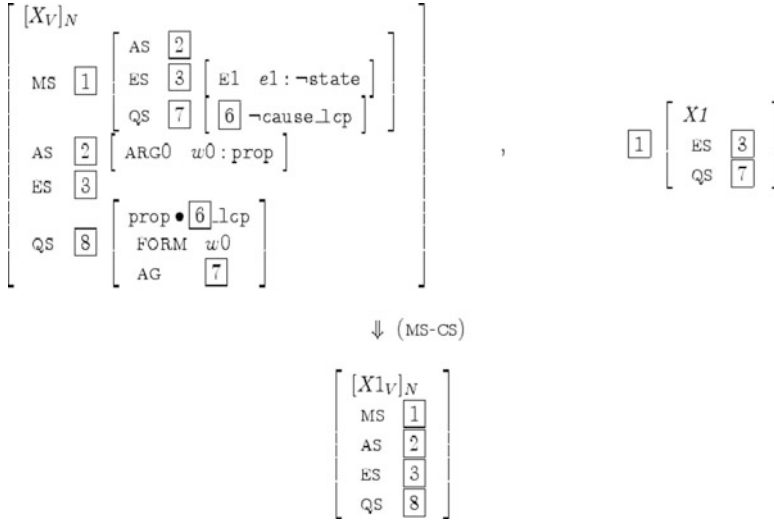


Fig. 17.8 MS-CS with VtoN conversion

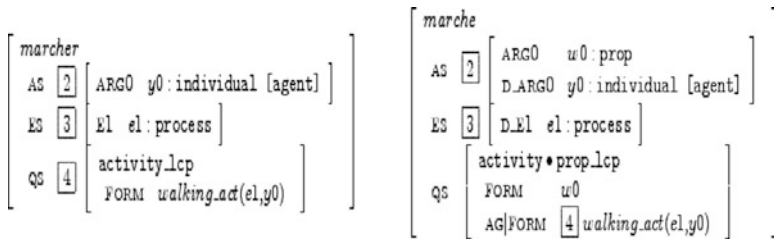


Fig. 17.9 VtoN conversion:  $\text{march}(er)_V \rightarrow_{\text{conv}} \text{marche}_N$

### 17.4.2.3 Examples

In the following section we illustrate the various verbal lexical types which can unify with MS in  $[X_V]_N$ ALU. The section starts with non causative verb types, i.e. the case of an activity input verb (*march(er)* [*walk<sub>V</sub>*]), followed by the case of a non-causative accomplishment (*transport(er)* [*carry<sub>V</sub>*]). Then, the last two examples are meant to indicate how the mechanism works in order to deal with input verbs which bear a causative interpretation (*recul(er)*, *angoiss(er)* [*distress<sub>V</sub>*]).

#### Activities

In simple activity (non telic) qualia structures, such as in *march(er)* (or *dans(er)* [*dance<sub>V</sub>*] or *chant(er)* [*sing<sub>V</sub>*] . . .), only the FORMAL role is defined, as shown on the lefthand side of Fig. 17.9. The successful MS-CS output deverbal noun appears

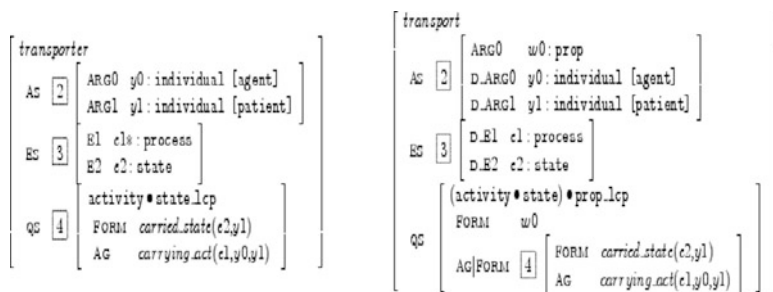


Fig. 17.10 VtoN conversion:  $transport(er)_V \rightarrow_{conv} transport_N$

on its righthand side. As indicated in Fig. 17.8, both input argument structure (MS|AS) and event structure (MS|ES) are inherited by the output noun, through, respectively, indices [2] and [3]. As for the input qualia structure (index [7] in Fig. 17.8, and [4] in Fig. 17.9), it matches against the noun AGENTIVE qualia value, according to  $[X_N]_V$  recommendations.  $[X_N]_V$  also imposes to the output noun an exocentric dotted type labelled with activity• proposition.lcp. This type results (1) from the successful unification of *march(er)* lexical entry with the  $[X_N]_V$ MS value (Fig. 17.8, index [1]), which means: (a) event types compatibility (*march(er)* does not designate a state), and (b) LCP compatibility (activity is a case of  $\neg$ cause); and (2) from the successful verb LCP propagation into the nominal LCP labelling (Fig. 17.8, index [6]). MS value [1] is propagated onto the output noun structure, though this is not represented in (Fig. 17.9). As shown by its QS, the exocentric dotted typed noun *marche* holds two readings: the first refers to a process *la marche des soldats sur la ville* [The soldiers' march on the city] or to its result: *trois longues marches en forêt* [three long walks in the forest] (depending on the agent **y0** realisation) and the second to a concept *la marche est une discipline olympique*, [race walk is an Olympic sport] activating only the QS|FORM value.

### Accomplishments

With *transport(er)*, we intend to illustrate (Fig. 17.10) a case of non-causative accomplishment. The input verb's event structure, headed by the process, is propagated into the output noun [3], together with its argument structure [2]. The MS-CS unification principle works in the same way as for *marche*, and gives raise to the definition of an exocentric dotted typed qualia structure composed with two mutually exclusive types: (prop) *un transport nécessite toujours un transporteur* [transports always require conveyors], and accomplishment, with type activity• state. Accomplishments can be realised, as nominal lexical units, either through the QS|AG|AG activity value: *le transport, lundi prochain, de la marchandise par le premier convoi* [Goods conveying, next Monday, by the first train], or through

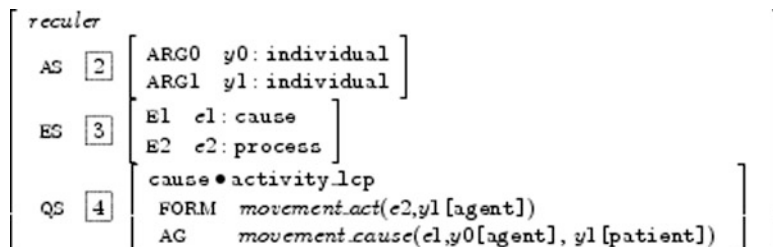


Fig. 17.11 Lexical entry of *recul(er)*<sub>v</sub>

the QS|AG|FORM resultative (state) value: *Tous les transports sont annulés jusqu'à lundi prochain*<sup>18</sup> [all transports are cancelled until next Monday].

Causatives

Let us now turn to more complex verb types or so it seems. The verbs *recul(er)* and *angoiss(er)* illustrate the case of causative predicates, that are mainly movement or psychological predicates. In *Max recule la chaise* [Max moves back the chair], the agent *Max* voluntarily causes the chair movement, and in *Le film a angoissé Max* [The movie distressed Max], the movie content entails *Max* psychological state of anxiety. These verbs generally carry a second resulting and intransitive reading. The subsequent movement for verb types like *recul(er)*: *Les ennemis reculent* [Enemies are going back], and the caused state for verb types like *angoiss(er)*: *Max angoisse* [Max is worried sick]. As J. Meinshäfer pointed out, only the non causative reading is an available candidate input for VtoN conversion: *\*le recul de la chaise par Max* [Max's moving back of the chair], versus *le recul de Max* [Max's backward movement], *\*l'angoisse de Max par le film* [The movie distress of Max], versus *l'angoisse de Max* [Max's distress].

From a formal point of view, these distinct, and non-overlapping verb interpretations are represented by exocentric dotted typed structures. As illustrated by (Fig. 17.11) *recul(er)* and by (Fig. 17.12) *angoiss(er)*, the activation of causative readings (AGENTIVE role) and that of resultative readings (FORMAL role) are therefore mutually exclusive.

Unifying *recul(er)*, as illustrated above, with the  $[X_N]_{VALU}$  MS (see Fig. 17.6) is above all in this case a matter of QUALIA types unification. In fact, the cause component within the cause • activity\_lcp labelled verbal exocentric dotted type is neutralised through unification with the  $\neg$ cause\_lcp required input verb. As this

<sup>18</sup>Since nouns carry only default event parameters, there is no headedness involved in their event structure, so, for accomplishments such as *transport*, both process and state can be realized. And since deverbal nouns inherit input verb argument structure content only as default arguments, noun arguments are always optional.



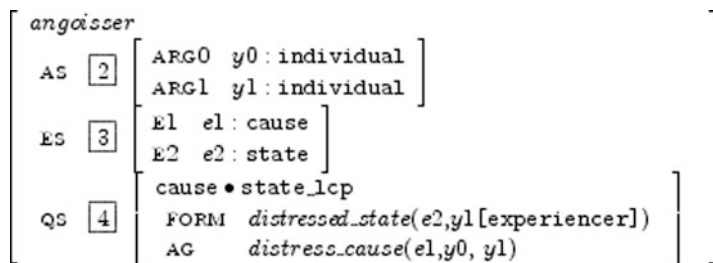


Fig. 17.12 Lexical entry of *angoisser*(er)<sub>v</sub>

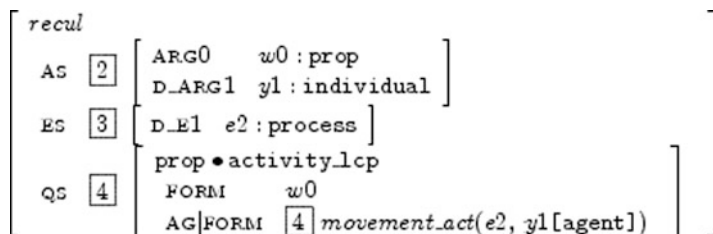


Fig. 17.13 VtoN conversion output *recul*<sub>N</sub>

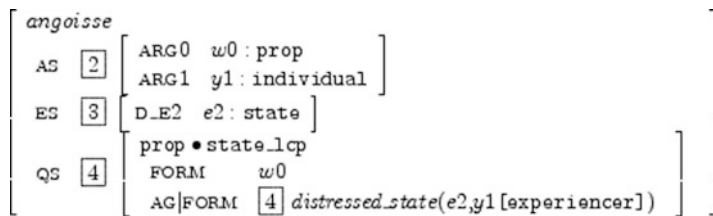


Fig. 17.14 VtoN conversion output *angoisse*<sub>N</sub>

component is filtered out, so are the corresponding qualia roles, together with their event and argument parameters. The unification effect is that of selecting only the verb resultative reading. The same occurs for *angoisser*(er): through unification with ALU MS, the causative interpretation is rejected, whereas the resultative static predicate is kept as the actual VtoN conversion input.

Once the correct reading has been selected, the remaining WF mechanism works on in a straightforward way: (1) the appropriate qualia label fulfills the missing slot on the output noun LCP (providing thus *recul* with proposition • activity\_lcp, and *angoisse* with proposition • state\_lcp), (2) the qualia structure defining the resultative predicate is inherited by the noun AGENTIVE value, while its FORMAL value is the prop typed index w0. Output nouns *recul* [*backward movement<sub>N</sub>/retreat<sub>N</sub>*] and *angoisse* [*distress<sub>N</sub>*] are displayed respectively in Figs. 17.13 and 17.14. It can be noticed that (1) *recul* may denote an agentive intransitive movement process (*le recul de l'armée* [*The army's retreat*]), or the movement result (*les reculs sont*

*inévitables* [Backward movements are unavoidable]), or a concept (*le recul s'oppose à l'avancée* [Backward movements are opposed to advancements]); (2) similarly, *angoisse* is a static nominal, that may or not affect an experiencer (*l'angoisse (de Max) a été provoquée par un stress* [(Max's) distress has been caused by stress]) or depict a concept (*l'angoisse est étudiée en psychanalyse* [Anguish is studied in psychology]).

## 17.5 Conclusion and Perspectives

In this paper, we have described a GL-based model designed for Word Formation. This model includes a composition schema called MS-CS and several abstract lexical units, each of which simulating a Word Formation process. This device has first been used to represent noun to verb *é*-prefixation in French (Jacquy and Namer to appear; Namer and Jacquy 2003). The robustness of the chosen approach has been confirmed when applied in the conversion framework presented here. The success of this approach is due to the fact that it combines linguistic hypotheses from a well-established morphological theory (inspired from D. Corbin work), as well as a lexical semantics formalism, namely GL (Generative Lexicon).

Moreover, coupling Word Formation theory with lexical semantics, through this method, has two additional effects:

- It makes obvious differences between seemingly identical phenomena. This paper has illustrated the structure distinctions for verbs such as *walk* and *dance*, or *drain* and *butter*, whereas they were analyzed in the same way in Pustejovsky (1996),<sup>19</sup> although they belong to opposite Word Formation families according to the morphological theory we rely on.
- It draws out similarities concealed behind apparent differences. Hence this paper has shown that verbs *drain(er)* and *sing(er)* result from a single WF rule, via  $[X_N]_{VALU}$ . Each time, only one mechanism is at play, their corresponding input nouns being responsible for the differences in verbal meanings.

Both similarities and differences are detected and analyzed within morphological theory; GL collects, ranks and formally expresses all of these linguistic hypotheses. In addition to this new collaboration between these two linguistic fields, the model also seems to provide new future prospects in Natural Language Processing. In fact, formalizing both NtoV and VtoN conversion, on the basis of a corpus analysis, can be viewed as an empirical checking of linguistic predictions about neologisms. In this regard, this experiment has confirmed the productivity of verb-to-noun (VtoN) conversion leading to processive nouns, or that of noun-to-verb (NtoV) conversion process leading to instrumental, locative or stereotypical agentive verbs. On the

---

<sup>19</sup>Actually, only *dance* and *butter* are mentioned in Pustejovsky (1996): we assume that *walk* and *drain* would be represented in the same way.

other hand, it has also allowed to detect the emergence of new, quantitatively important classes: nominal verbs denoting a change-of-state (*marbr(er)*, *guêp(er)*), borrowings and onomatopoeias Noun/Verb pairs denoting (speech) acts or sounds (*patch/patch(er)*, *blabla/blablat(er)*, *glouglou/glouglout(er)*).

Identifying the most creative conversion types, predicting their semantic constraints exerted on both input and output, and drawing their input-to-output semantic relationships, through the choice of the right conversion orientation are results which could be used further in NLP systems in order to enrich lexical contents.

## References

- Adams, V. (1973). *An introduction to modern English word-formation*. London: Longman Group Limited.
- Anschombe, J.-C. (1979). Délocutivité benvenistienne, délocutivité généralisée et performativité. *Langue Française*, 42, 69–84, Paris.
- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Aronoff, M., Oehrle, R., Kelley, F., & Stephens, B. W. (Eds.). (1984). *Language sound and structure*. Cambridge, MA: MIT Press.
- Aurnague, M., & Plénat, M. (1996). La préfixation en é- et la relation de partie à tout. In D. Genthial et al. (Eds.), *Seminaire Lexique. Représentations et outils pour les bases lexicales. Morphologie robuste* (pp. 43–52). Grenoble: CLIPS-IMAG, Université de Grenoble.
- Aurnague, M., & Plénat, M. (1997). Manifestations morphologiques de la relation d'attachement habituel. In D. Corbin, G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat, & M. Roché (Eds.), *Silexicales 1: Mots possibles, mots existants* (pp. 15–24). Villeneuve d'Ascq: Presses de l'Université de Lille III.
- Benveniste, E. (1966). *Problèmes de linguistique générale*. Paris: Gallimard.
- Bybee, J. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology* (pp. 119–141). San Diego: Academic.
- Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique* (2 Vols.). Max Niemeyer Verlag, Tübingen (2nd ed.). Villeneuve d'Ascq: Presses Universitaires de Lille.
- Corbin, D. (2001). Préfixes et suffixes: du sens aux catégories. *Faits de langue*, 15, 41–69, Ophrys, Paris.
- Corbin, D. (2004). French (Indo-European: Romance). In G. Booij, C. Lehmann, & J. Mugdan (Eds.), *Morphology. An international handbook on inflection and word formation* (Vol. 1). New York: Walter de Gruyter, art. 121.
- Cornulier, B. d. (1976). La notion de dérivation délocutive. *Revue de linguistique romane*, 40, 116–144, Champion, Paris.
- Cruse, D. A. (1986). *Lexical semantics*. London: Cambridge University Press.
- Dal, G. (1997). Du principe d'unicité catégorielle au principe d'unicité sémantique: incidence sur la formalisation du lexique construit morphologiquement. In P. Greenfield, H. Madec, P.-A. Buvet, & S. Cardy (Eds.), *Actes de Fractal'97, BULAG numéro spécial* (pp. 105–115). Besançon: Presses universitaires de Franche-Comté.
- Di Sciullo, A.-M., & Williams, E. (1987). *On the definition of word*. Cambridge: MIT Press.
- Fradin, B. (2003). *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.
- Fradin, B., & Kerleroux, F. (2003a). Quelles bases pour les procédés de la morphologie constructionnelle ? In B. Fradin, G. Dal, N. Hathout, F. Kerleroux, M. Plénat, & M. Roché (Eds.), *Silexicales 3: Les unités morphologiques* (pp. 76–84). Villeneuve d'Ascq: SILEX: CNRS & Université de Lille 3.

- Fradin, B., & Kerleroux, F. (2003b). Troubles with lexemes. In G. Booij, J. De Cesaris, A. Ralli, & S. Scalise (Eds.), *Topics in morphology. Selected papers from the third mediterranean morphology meeting* (pp. 177–196). Barcelona: Institut Universitari de lingüística aplicada, Univeritat Pompeu Fabra.
- Hale, K., & Kayser, S. J. (1993). On argument structure and the lexical representation of syntactic relations. In K. Hale & S. J. Kayser (Eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 53–110). Cambridge, MA/Cambridge, UK: MIT Press.
- Haspelmath, M. (2002). *Understanding morphology*. London: Arnold.
- Jacquy, E., & Namer, F. (2003). Morphosémantique et modélisation: les verbes dénominaux préfixés par é-.in: Actes de la seconde conférence “Représentation du sens linguistique”, Montréal.
- Katamba, F. (1993). *Morphology*. Houndmills, Basingstoke: Palgrave Macmillan.
- Kerleroux, F. (1996a). *La coupure invisible: études de syntaxe et de morphologie*. Villeneuve d’Ascq: Presses universitaires du Septentrion.
- Kerleroux, F. (1996b). Représentations de l’absence de suffixe dans les noms déverbaux processifs du français. In *Travaux linguistiques du CERLICO, n°9* (pp. 141–170). Rennes: Presses Universitaires de Rennes.
- Kerleroux, F. (1997). De la limitation de l’homonymie entre noms déverbaux convertis et apocopes de noms déverbaux suffixés. In D. Corbin, G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat, & M. Roché (Eds.), *Silexicales 1: Mots possibles, mots existants* (pp. 163–172). Villeneuve d’Ascq: Presses de l’Université de Lille III.
- Kerleroux, F. (1999). Identification d’un procédé morphologique: la conversion. *Faits de langues, 14*, 89–100. Ophrys, Paris.
- Kerleroux, F. (2004). Sur quels objets portent les opérations morphologiques de construction ? In *Lexique 16* (pp. 85–124). Villeneuve d’Ascq: Presses universitaires du Septentrion.
- Kiparsky, P. (1982). Lexical phonology and morphology. In I.-S. Yang (Ed.), *Linguistics in the morning calm* (pp. 3–91). Seoul: Hanshin.
- Koenig, J.-P. (1999). *Lexical relations*. Stanford: CSLI publications.
- Larcher, P. (2003). La dérivation délocutive: Histoire d’une notion méconnue. *Historiographia Linguistica, 30*(3), 389–406(8), John Benjamins Publishing Company, Amsterdam.
- Lieber, R. (1992). *Deconstructing morphology*. Chicago: University of Chicago Press.
- Marchand, H. (1969). *The categories and types of present-day English word-formation*. Munich: C.H. Beck Verlagsbuchhandlung.
- Meinschäfer, J. (2003). Remarques sur l’interprétation des déverbaux sans affixe en français. In B. Fradin, G. Dal, F. Kerleroux, N. Hathout, M. Plénat, & M. Roché (Eds.), *Silexicales 3: Les unités morphologiques* (pp. 118–125). Villeneuve d’Ascq: Presses universitaires de Lille.
- Mel’cuk, I. (1996). *Cours de morphologie générale* (Vol. 3). Montréal/Paris: Presses de l’université de Montréal -CNRS.
- Mel’cuk, I. (1997). *Cours de morphologie générale* (Vol. 4). Montréal/Paris: Presses de l’université de Montréal -CNRS.
- Namer, F. (2003). Automatiser l’analyse morpho-sémantique non affixale: le système DériF. In *Cahiers de Grammaire 28* (pp. 31–48). Toulouse: Presses universitaires du Mirail.
- Namer, F., & Jacquy, E. (2003). Lexical semantics and derivational morphology: The case of the popular ‘é-’ prefixation in French. *Proceedings of the Second Workshop on Generative Approaches to the Lexicon, Geneva*, pp. 115–122.
- Plénat, M., & Roché, M. (2004). Prosodic constraints on suffixation in French. In G. Booij, J. De Cesaris, A. Ralli, & S. Scalise (Eds.), *Topics in morphology. Selected papers from the third Mediterranean morphology meeting* (pp. 285–299). Barcelona: Institut Universitari de lingüística aplicada, Univeritat Pompeu Fabra.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Pustejovsky, J. (1996). *Lexical shadowing and argument closure, ms*. Boston: Brandeis University.

- Sanders, G. (1988). Zero derivation and the overt analogue criterion. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: Approaches in modern linguistics* (pp. 155–175). San Diego: Academic.
- Selkirk, E. (1982). *The syntax of words*, *Linguistic Inquiry Monograph*. Cambridge, MA: MIT Press.
- Shieber, S. M. (1986). *An introduction to unification-based approaches to grammar* (CSLI Lecture Notes Series, Vol. 4). Stanford: Center for the Study of Language and Information.