French VN lexemes: morphological compounding in HPSG

Florence Villoing, Marianne Desmets

To cite this version:


HAL Id: halshs-00530138
https://halshs.archives-ouvertes.fr/halshs-00530138
Submitted on 28 Oct 2010

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
French VN lexemes: morphological compounding in HPSG

Marianne Desmets           Florence Villoing
U. Paris Ouest/LLF         U. Paris Saint-Denis/SFL

Proceedings of the HPSG09 Conference

Georg-August-University at Göttingen, Germany

Stefan Müller (Editor)

2009

CSLI Publications

http://csli-publications.stanford.edu/
Abstract

Although the original framework of HPSG is mostly compatible with independent theoretical claims or analyses in lexeme base morphology (Anderson 1992, Aronoff & Fudeman 2004, Beard 1995, Booij 2005, Carstairs-McCarthy 1992, Fradin 2003, Haspelmath 2002, Matthews 1991, Plag 2003, for example), so far, most morphological research in morphology has been done on inflexional phenomena (Orgun & Inkelas 2002, Bonami & Boyé 2006), and few on derivational morphology has been addressed by only a few (Koenig 1999, Riehemann 1998). Yet, we believe it is worth investigating how the formal and theoretical apparatus of HPSG deals with capturing multilevel constraints that apply in the lexeme formation of French Verb-Noun nominal compounds in French (, such as as GRILLE-PAIN, (lit. grill-bread, ‘toaster’), PERCE-OREILLE, (lit. pierce-ear, ‘earwig’), TOURNEVIS, (lit. turn-screw, ‘screwdriver’), or LÈCHE-VITRINE, (lit. lick-window, ‘window-shopping’), can be captured by the formal and theoretical apparatus of HPSG. Contrary to the view what has often expressed in the past been said, we argue that VN lexemes formation comes under is subject to morphological constraints rather than to but not under syntactic mechanisms. Our analysis integrates VN lexemes into a multiple-dimensional typed-hierarchy of lexemes and provides an account for of semantic generalizations involved in different types of lexeme formation (compounding, derivation, and conversion).

Morphological compounding is a mechanism of lexeme formation that has been studied less within HPSG compared to derivational and inflexional phenomena. In this paper, we propose a morphological treatment of French Verb-Noun compounds (as in 1), which have been frequently considered as lexicalized syntactic phrases in the literature. We present an HPSG analysis that integrates compounding in a general lexeme typed-hierarchy, and captures some generalities about the semantics needed in most deverbal lexeme formations, in particular, in VN lexemes, derived lexemes, and convert lexemes.

(1)  
   a. GRILLE-PAIN\(^2\) (lit. grill-bread, ‘toaster’)  
   b. PERCE-OREILLE (lit. pierce-ear, ‘earwig’)  
   c. TOURNEVIS (lit. turn-screw, ‘screwdriver’)  
   d. LÈCHE-VITRINE (lit. lick-window ‘window-shopping’)

1 Why VN compounds are not syntactic formations

In the literature, French VN compounds are commonly considered as syntactic formations (Di-Sciullo & Williams 1987, Barbaud 1994, Lieber 1992, Zwanenburg 1992, among others). However, following Corbin (1992), Fradin (2005) and Villoing (2009), we argue that VN compounds do not

---

1 We would like to thank the members of the ‘Lectures in HPSG’ seminar at Paris-Diderot University, specially Anne Abeillé, Olivier Bonami, Danièle Godard, Bernard Fradin and Françoise Kerleroux for their remarks and substantive suggestions; we would also like to thank the members of the HPSG’09 conference for their interesting questions.
2 By convention, lexemes are in small capitals.
show the properties expected of lexicalized syntactic phrases, a fact which argues in favor of the idea that they are formed morphologically rather than syntactically.

First, compounding as the morphological formation of lexemes does not typically involve functional words. Determiners, prepositions, and pronouns (including clitic forms, which are inflexional forms, see Miller & Sag 1997) never realize in a compound. Remarkably, the type of nouns selected by the VN compound rule always appears with a determiner in the corresponding sentence (cf. 2), while determiners never realize with the N in VN nominals, cf. (3).

(2) Cet objet grille le pain./*Cet objet grille pain. 
it. this object grill the bread. / this object grill bread
(3) GRILLE-PAINN (lit. grill-bread, ‘toaster’)
Cet objet est un grille-pain / *Cet objet est un grille-le-pain.
it. this object is a grill-bread. / *This object is a grill the bread

In contrast, syntactic lexicalization of verb phrases, including those involving the same categories (V and N) as VN compounds (cf. 5), do characteristically preserve functional words of the original syntactic phrase, including prepositions (4a), pronouns (4b,c), and determiners (5):

(4) a. BOIT-SANS-SOIFN (lit. drinks-without-thirst, ‘drunkard’)
b. RENDEZ-VOUSN (lit. go-you, ‘appointment’) 
c. SOT-L’Y-LAISSEN (lit. silly-it-there-leaves, ‘chicken oyster’)
(5) a. TROMPE L’OEILN (lit. deceives-the-eye, ‘trompe l’oeil’)
b. TROMPE-LA-MORTN (lit. deceives-the-death, ‘daredevil’)
c. CRÈVE-LA-FAIMN (lit.dies-the-hunger,’beggar, destitute person’) 
d. PUE-LA-SUEURN (lit. stinks the perspiration, ‘poor laborer’)

In addition, if VN compounds were lexicalized phrases, inherent reflexive pronouns that are obligatorily realized in the syntax would be expected to appear. However, this is not the case, as demonstrated in (6):

(6) a. Il se casse la tête. (lit. he reflex breaks the head)
b. C’est un casse-tête. (lit. it’s a break-head)
c. *C’est un se casse-tête (lit. it’s a reflex break-head)

Second, lexicalized phrases preserve in their structure the original SVO word order of the source sentence, as in (7). In light of this property, it is interesting to observe that many Verb-Noun compounds cannot appear as such in a sentence, specifically because the N realized on the right of the verb does not satisfy the syntactic constraints on the realization of the semantic arguments of the verb. The N of the VN lexemes in (8), for example, is understood as an agent and would be realized in a sentence as a subject on the left of the verb.
In lexeme compounds, the semantic relations between the verb and noun is not absolutely uniform, nor as predictable as it would be in a syntactic structure. While the N in a VN compound most frequently denotes the patient of the verb (cf. 9a), it can also denote other roles, such as the agent (cf. 8), spatial localization (cf.9b), or temporal localization (cf. 9c):

(9)  
\[ \begin{align*}
\text{a. OUVRE-BOÎTE}_N & \quad \text{(lit. open-tin, ‘tin opener’)} \\
\text{b. TRAÎNE-BUSSON}_N & \quad \text{(lit. hang around on-bush, ‘animal’)} \\
\text{c. REVEILLE-MATIN}_N & \quad \text{(it. wake up-morning, ‘alarm clock’)}
\end{align*} \]

In fact, this relative plasticity of argument interpretation is a characteristic of lexeme compounding, and contrasts with the limited range of interpretation exhibited by the argument structure of a given verb in a sentence. As another illustration of this phenomenon, we observe that the resulting interpretation of a VN compound may also vary, even for a given verb-noun semantic relation. For example, among the patient relations in (10), VN (10a) denotes a patient, VN (10b) denotes an event, and VN (10c) an agent (and others may denote an instrument, or a localization).

(10)  
\[ \begin{align*}
\text{a. GOBE-MOUTON}_N & \quad \text{(lit. swallow-sheep, ‘poisonous plant’)} \\
\text{b. LECHE-VITRINE}_N & \quad \text{(lit. lick-window, ‘window shopping’)} \\
\text{c. GRATTE-PAPIER}_N & \quad \text{(lit. scratch-paper, ‘pen pusher’)}
\end{align*} \]

As VN compounds do not exhibit syntactic constraints that are preserved in lexicalized phrases, we conclude that these compounds are morphological constructs in French (Corbin 1992, Villoing 2003, Fradin 2005).

2 The lexeme properties of VN compounds

2.1 General properties

Verb-Noun compounds are nominals. As morphological constructs, they are formed of two lexemes: a verbal base-lexeme and a nominal base-lexeme. The semantics of the whole compound (S3, in Table 1) involves the semantics of the base-lexemes AND the semantics of the morphological rule.
\[ F1: x \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \ quad
b. Le gril-

le gras est cassé. (Subj: Spec+N)

tl. The grill-

brea

d ‘toaster’) is broken.

The same properties are observed for event-denoting VN

s, as in (17a-c):

(17) a. Le lêche-

vitri
c est mon loisir préféré (Subj: Spec+N)

lit. the window-shopping is my favorite hobby

b. Marie adore le lêche-

vitri
c (Obj: Spec+N)

lit. Mary loves window-shopping

c. Lola fait du lêche-

vitri
c (Obj : Spec indef +N)

lit. Lola goes window-shopping

The semantics of the compounding rule allow object-denoting VN

s to be

predicative (18a, b) or attributive (18c):

(18) a. Paul a acheté du papier tue-

mouche.

lit. Paul bought some kill-fly paper (‘flypaper’)

b. Ce couloir a trois portes coupe-

feu.

lit. This corridor has three cut-fire doors (‘firebreak door’)

c. Pierre est rabat-

joie.

lit. Peter is reduce-joy (‘spoilsport’).

In this case, the modified N (papier in (18a), portes in (18b)) or the subject

(Pierre in (18c)) is the Proto-agent of the verbal base-lexeme (in the sense of

dowty 1991). In the predicative use, the paper is seen as the killer of flies

(18a) and the door as the one that cuts fire (18b). In the attributive use,

Pierre, a human, is seen as the one who causes the reduction of joy (18c).

Event-denoting VN

s can also have attributive or predicative uses, since

French allows the construction of VN

es denoting a property from an event, cf.

(19) and (20). But, this is neither direct nor systematic, and requires some

semantic accommodation.

(19) a. On part pour une journée lêche-

vitri
c. (web)

lit. We are going for a day window-shopping

(20) Il est très bais-

main.

lit. He is very kiss-hand

2.4 The phonological properties of VN compounds

Many discussions have focused on the nature of the verb in VN compounds,

especially on the question whether it is a stem or a word-form (see Villoing

1999 for an overview). Since VN verb forms are not marked for inflection,

we consider them stems. As for their phonological properties, we follow
Bonami & Boyé’s (2003) account of verbal inflection in French. In their perspective (following, among others, Aronoff 1994 for Latin conjugation), verbal lexemes are associated in the lexicon with a vector of different possible phonological representations. These phonological representations are distinct stems, which Bonami & Boyé (2003) call the “stem space”.

<table>
<thead>
<tr>
<th>Lexeme</th>
<th>Stem 1 (PRST. SG)</th>
<th>Stem 2 (PRST.3.PL)</th>
<th>Stem 3 (PRST. 1/2 PL IMPARF.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOIRE ‘to drink’</td>
<td>/bwa/</td>
<td>/bwav/</td>
<td>/byv/</td>
</tr>
</tbody>
</table>

Table (2): The phonological verb stem of the verb BOIRE

Each lexical or inflectional morphological rule selects for a specific stem as input. From the possible stems of the verb, the VN compounding rule always selects for stem 1. The verb lexeme SOUTENIR, for example, has at least two stems /sut ɔ̃/ and /sutʒɛ/; the rule selects for the first, which is also used to form the present singular.

<table>
<thead>
<tr>
<th>Verb lexeme</th>
<th>Stem 1 (PRST SG)</th>
<th>VN compound</th>
</tr>
</thead>
<tbody>
<tr>
<td>COUPER ‘to cut’</td>
<td>/k ɔ̃p/</td>
<td>COUPE-PAPIER_\textsubscript{N} lit. cut-paper, ‘paper knife’</td>
</tr>
<tr>
<td>LECHER ‘to lick’</td>
<td>/lɛʃ/</td>
<td>LÊCHE-VITRINE\textsubscript{N} lit.lick-window, ‘window-shopping’</td>
</tr>
<tr>
<td>ESSUYER ‘to wipe’</td>
<td>/ɛsɥi/</td>
<td>ESSUIE-GLACE\textsubscript{N} lit.wipe-window, ‘windshield wiper’</td>
</tr>
<tr>
<td>OUVRIR ‘to open’</td>
<td>/u ɔ̃v/</td>
<td>OUVRRE-BOÎTE\textsubscript{N} lit. open-tin, ‘tin opener’</td>
</tr>
<tr>
<td>SOUTENIR ‘to support’</td>
<td>/sutʃɛ/</td>
<td>SOUTIEN-GORGE\textsubscript{N} lit.support-bosom, ‘bra’</td>
</tr>
<tr>
<td>TORDRE ‘to wring’</td>
<td>/tɔʁ/</td>
<td>TORD-BOYAU\textsubscript{N} lit.wring-gut, ‘rotgut’</td>
</tr>
</tbody>
</table>

Table (3): The phonological verb stem of VN compounds

The first stem is the default phonological stem for all verbs involved in the VN compounding rule, while the default stem for derivational rules is commonly stem 3, used for the present plural or for the perfect tense.
<table>
<thead>
<tr>
<th>Verb lexeme</th>
<th>Stem 1 (PRST SG)</th>
<th>VN compound</th>
<th>Stem 3 (PERFECT)</th>
<th>Deverbal nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESSUYER ‘to wipe’</td>
<td>/esˈyi/</td>
<td>ESSUIE-GLACE\textsubscript{N}</td>
<td>/esˈyi/</td>
<td>ESSUYAGE ‘drying up’ ESSUYEUR ‘dryer’</td>
</tr>
<tr>
<td>SOUTENIR ‘to support’</td>
<td>/sutˈje/</td>
<td>SOUTIEN-GORGE\textsubscript{N}</td>
<td>/sutən/</td>
<td>SOUTENABLE ‘bearable’ SOUTENANCE ‘academic defense’</td>
</tr>
<tr>
<td>TORDRE ‘to wring’</td>
<td>/tɔʁ/</td>
<td>TORD-BOYAUN</td>
<td>/tɔrd/</td>
<td>TORDABLE ‘wringable’ TORDEUR ‘wringer’</td>
</tr>
</tbody>
</table>

Table (4): The phonological verb stem of VN compounds and deverbal nouns

The noun can, in most cases, be analyzed as a stem, but may sometimes look like a word form marked for plural:

(21) a. ESSUIE-MAINS\textsubscript{N} (lit. dry-hands, ‘hand towel’)  
b. PRESSE-FRUTS\textsubscript{N} (lit. press-fruits, ‘squeeze’)  
c. PROTÈGE-YEUX\textsubscript{N} (lit. protect-eyes, ‘eye mask’)  

We believe that this is not syntactic marking, but an inherent inflection (such as described by Booij, 1996). Inherent inflection is required by the semantics and not by the syntax. The choice of singular or plural marking by the rule does not really change the semantics of the whole VN.

2.5 The semantic properties of VN

2.5.1 The semantics of the verbal base-lexeme

The verbal base-lexeme of a VN is dynamic (following Vendler 1967 and Dowty 1979). Stative verbs are, therefore, bad candidates for VN compounding:

(22) a. ?? Paul est un véritable sait-latin.  
    lit. Paul is a true know-Latin  
    b. ?? Le Béluga, les aime-caviar russes en sont fous.  
    approx. The Beluga, Russian love-caviars are crazy about it

Most verb bases are transitive and present an agent/patient relation. Therefore, unaccusative verbs (23) and unergative verbs (24) are typically bad candidates as well:

(23) a. ?? °ARRIVE-TRAIN\textsubscript{N} (lit. arrive-train)  
b. ?? °TOMBE-PLUIE\textsubscript{N} (lit. fall-rain)
Nevertheless, some VN compounds are built on unaccusative or unergative verb bases; in this case, the interpretation is causative, through an agent participant variable added by the rule (see 37-40 below).

### 2.5.2 The semantics of the nominal base-lexeme

In most cases, the semantic role of the noun-base is the Proto-patient argument of the verbal base-lexeme, as in (25). So, the noun base denotes what is affected by the process described by the verb. In rare cases, it may be understood as another argument: agent, spatial or temporal localization. All the possibilities are summed up in Table (5).

(25)

<table>
<thead>
<tr>
<th>Compound</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LÈCHE-VITRINE&lt;sub&gt;N&lt;/sub&gt;</td>
<td>(lit. lick-window, ‘window-shopping’)</td>
</tr>
<tr>
<td>OUVRE-BOÎTE&lt;sub&gt;N&lt;/sub&gt;</td>
<td>(lit. open-tin, ‘tin opener’)</td>
</tr>
<tr>
<td>GRATTE-PAPIER&lt;sub&gt;N&lt;/sub&gt;</td>
<td>(lit. scratch-paper, ‘pen pusher’)</td>
</tr>
<tr>
<td>COUPE-GORGE&lt;sub&gt;N&lt;/sub&gt;</td>
<td>(lit. cut-throat, ‘dangerous back alley’)</td>
</tr>
<tr>
<td>TROTTE-BÉBÉ&lt;sub&gt;N&lt;/sub&gt;</td>
<td>(lit. toddle-along-baby, ‘baby walker’)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Agent</th>
<th>Location</th>
<th>Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>LÈCHE-VITRINE</td>
<td>GOBE-MOUTON</td>
<td>TRAÎNE-BUISSON</td>
<td>RÉVEILLE-MATIN</td>
</tr>
<tr>
<td></td>
<td>HURLE-LOUP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PISSE-VACHE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TROTTE-BÉBÉ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUVRE-BOÎTE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRATTE-PAPIER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COUPE-GORGE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table (5): The semantic role of the N in a VN compound

### 2.5.3 Semantic properties of the whole VN

As we said above, VN compounding has two possible types of semantic output: event-denoting nominals and object-denoting nominals. Event-denoting nominals, as in (26), denote a subset of events:

(26)

<table>
<thead>
<tr>
<th>Compound</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LÈCHE-VITRINE&lt;sub&gt;N&lt;/sub&gt;</td>
<td>(lit. lick-window, ‘window-shopping’)</td>
</tr>
<tr>
<td>REMUE-MENAGE&lt;sub&gt;N&lt;/sub&gt;</td>
<td>(lit. move-household, ‘commotion’)</td>
</tr>
</tbody>
</table>

---

<sup>3</sup> The Proto-patient, as well as the Proto-agent, are defined according to the criteria given by Dowty (1991) and Davis & Koenig (2000).

<sup>4</sup> Boldface indicates the most common interpretation.
Object-denoting nominals are obtained from two different types of semantic rules.

• The first rule operates the abstraction of a participant variable of the verbal base-lexeme. In this case, VN compounds are mostly interpreted as the Proto-agent, as in (27):

(27)  
  a. OUVRE-BOITE
       (lit. open-tin, ‘tin opener’)  
  b. REVEILLE-MATIN
       (lit.wake up-morning, ‘alarm clock’)  
  c. GRATTE-PAPIER
       (lit. scratch-paper, ‘pen pusher’)  
  d. GARDE-COTE
       (lit. watch-coast, ‘coastguard’)

But, in a few other cases, it may also be a patient (cf. 28) or a spatial localization (cf. 29), as noted above. Table (6) sums up the various denotation types available for object-denoting VNs that correspond to the abstraction of a variable.

(28)  
  a. GOBE-MOUTON
       (lit. swallow-sheep, ‘kind of poisonous plant’)  
  b. BROUTE-BIQUET
       (lit. graze-kid (young goat), ‘honesuckle’)

(29)  
  a. COUPE-GORGE
       (lit. cut-throat ‘dangerous back alley’)  
  b. HURLE-LOUP
       (lit. howl-wolf, ‘toponym’)

<table>
<thead>
<tr>
<th>(proto)Agent</th>
<th>(proto)Patient</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRATTE-PAPIER</td>
<td>GOBE-MOUTON</td>
<td>HURLE-LOUP</td>
</tr>
<tr>
<td>TRAÎNE-SAVATE</td>
<td>BROUTE-BIQUET</td>
<td>COUPE-GORGE</td>
</tr>
<tr>
<td>OUVRE-BOÎTE</td>
<td></td>
<td>GARDE-MEUBLE.</td>
</tr>
<tr>
<td>REVEILLE-MATIN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table (6): The semantic role of VN compounds (select a participant)

• The second semantic rule involved in object-denoting VNs adds an agent participant variable to the verbal base-lexeme relation, via a causative relation in the case of a non-agentive verb base-lexeme (inaccusative), as in (30), or an instrumental relation in the case of unergative verbal bases, as in (31):

(30)  
  a. COULE-SANG
       (lit. flow-blood, ‘type of plant’)  
  b. SAUTE-BOUCHON
       (lit. jump-cork, ‘champagne’)

(31)  
  a. PISSE-CHIEN
       (lit. pee-dog, ‘type of plant’)  
  b. TROTTE-BÉBÉ
       (lit.toddle-along-baby, ‘baby walker’)

3 Analysis

3.1. A type hierarchy for morphologically complex lexemes

In Bonami & Boyer’s (2006) sign type hierarchy that we choose to use\(^5\), lexemic properties of a lex-sign (i.e., words and lexemes) are expressed via the attribute MORPHOLOGICAL-DAUGHTERS. This allows us to express that a word is a syntactic sign with a lexemic identity. This hierarchy also allows the distinction between words and lexeme signs, stipulating that PHONOLOGY is an attribute of syntactic signs (i.e. phrases and words), whereas the phonological identity of lexeme signs is expressed via the STEMS feature (see table (7) below).

![Diagram](image)

**Fig(1). Bonami & Boyé’s (2006) sign type hierarchy**

<table>
<thead>
<tr>
<th>TYPE</th>
<th>CONSTRAINT</th>
<th>ISA</th>
</tr>
</thead>
<tbody>
<tr>
<td>syn-sign</td>
<td>[PHON phon]</td>
<td>sign</td>
</tr>
<tr>
<td>lex-sign</td>
<td>[M-DTRS list(lexeme)]</td>
<td>sign</td>
</tr>
<tr>
<td>phrase</td>
<td>[ DTRS list(syn-sign)]</td>
<td>syn-sign</td>
</tr>
<tr>
<td>word</td>
<td>[M-DTRS &amp;lexeme&gt;</td>
<td>syn-sign &amp; lex-sign</td>
</tr>
<tr>
<td>lexeme</td>
<td>[STEMS stem-space]</td>
<td>lex-sign</td>
</tr>
</tbody>
</table>

**Table (7): Constraints on the sign type hierarchy**

The lexicon of languages builds lexemes by different means; this includes a widespread distinction (in French, as in other European languages) between simple lexemes (simplex) and morphologically complex ones. We propose to account for this variety of organization by using a further dimension of classification, called FORMATION, in addition to the PART-OF-SPEECH and VALENCE dimensions, see Fig. (2).

Lexemes with a complex morphology (morph-complex-lex) are classified into compound, derived and converted lexemes\(^6\). This analysis is based on several recent works in morphology; in particular, we integrate the results of Tribout (forthcoming) on converted lexemes, Fradin & Kerleroux (2002) and

\(^5\) Our analysis could also be mapped onto the SBCG framework (Sag, 2007), considering vn-lex as a type of construct.

\(^6\) We believe that inflected signs are syntactic-signs, hence, INFLECTION should be a dimension of word type hierarchy.

Fig (2). A multi-dimensional lexeme type hierarchy

3.2. Semantic rules available for deverbal lexeme formation

What emerges from these analyses is the fact that semantic rules involved in the formation of deverbal lexemes have much in common, whether these are compound, derived or converted lexemes. First, these always involve the semantic argument structure of the verb base. To account for this factor, we propose to use a type hierarchy for semantic roles à la Davis and Koenig (2000), as follow:

Fig (3). The role-relation (partial) type hierarchy

Second, two general semantic patterns are evident: deverbal lexemes may denote an event (or a set of events) or a referential index. The latter may be abstracted from the semantic argument structure of the verb base or be an additional argument. These general semantic patterns are captured in the complex-nominal-relation type hierarchy we propose:

Fig (4). The complex-nom-rel (partial) type hierarchy

Constraints on complex-nominal-relation, given in (32), are rather general, since each specific lexeme imposes its own particular semantics. Nominal-event-id-relation is an identity relation that takes an austinian as an
argument whose SITUATION index, corresponding to the index of the verb base, is identified with the INSTANCE value of the relation. This relation applies to event nominalizations in general: in addition to VN lexemes (like LÈCHE-VITRINE ‘window shopping’ or SAUTE-MOUTON ‘leapfrog’), it also applies to converted lexemes (NAGE ‘swimming’, CHUTE ‘downfall’), and to derived lexemes with the suffix -ette (BRONZETTE ‘sunbathing’), with the suffix -age (BALAYAGE ‘sweeping’) or with the suffix -on (PLONGEON ‘dive’).

Selectagent-rel, selectpatient-rel, and selectloc-rel are relations in which a particular semantic argument is abstracted from a verb base relation. It may denote an agent, as in the VN lexemes GRATTE-PAPIER (‘pen pusher’), GARDE-BARRIÈRE (‘gate keeper’), and in derived lexemes with the suffix -eur (MARCHEUR ‘walker’, CHANTEUR ‘singer’), with the suffix -on (GROGNON ‘grumbling’, BROUILLON ‘draft’), or in converted lexemes (JUGE ‘judge’, GARDE ‘guard’). It may denote a patient, as in the VN lexeme GOBE-MOUTON (‘kind of poisonous plant’), in derived lexemes with the suffix -oir (TIROIR ‘drawer’), with the suffix -ette (SUCETTE ‘lollypop’), with the suffix -on (NOURRISSON ‘infant’, SUÇON ‘hickey’), or in converted lexemes (AFFICHE ‘poster’, PARCOURS ‘route’). The abstracted semantic argument may also denote the localization of an event (or a set of events), as in the VN lexeme GARDE-MEUBLE (‘storage’), in derived lexemes with the suffix -oir (LAVOIR ‘wash house’, FUMOIR ‘smocking room’), in lexemes with the suffix -ette (BUVETTE ‘taproom’, CACHETTE ‘hiding place’), or in converted lexemes (INSTITUT ‘institute’, ARRIVÉE ‘arrival’).

The causal-relation and instrumental-relation are mostly used in cases where an ‘external’ agent is added to the argument structure of a verb base. The

---

7 We borrow the austinian type from Ginzburg and Sag (2000).
causal-relation adds a cause argument, mostly to a patient-only type of verb base, like the VN lexeme COULE-SANG (‘plant’, lit. leek–blood). The instrumental-relation adds an argument understood as an instrument or a mean to a verb base that has an agent-rel type of relation, like the VN lexeme TROTTE-BÉBÉ (‘baby walker’), or in derived lexemes with the suffix -oire (PASOIRE ‘strainer’, MACHOIRE ‘jaw’), with the suffix -eur (CHARGEUR ‘cartridge’, INTERRUPTEUR ‘switch’), with the suffix -ette (ALLUMETTE ‘matchstick’), or with the suffix -on (GUIDON ‘handlebar’, TORCHON ‘dishcloth’).

3.3. A type hierarchy for VN compounds

As expected, the first partition of the vn-lexeme type hierarchy we propose is between event-denoting nouns (nevent-vn-lex) and object denoting nouns (nobj-vn-lex). There are three subtypes of nobj-vn-lex: agent-vn-lex, patient-vn-lex and localization-vn-lex. Many VN lexemes are of the general agent-vn-lex type (that selects the agent argument of a transitive verb base), and some other agent-vn-lex are built by adding a cause or an instrument argument to the argument structure of the verb base.

![Fig (5). The vn-lexeme type hierarchy](image)

The following table lists these subtypes with examples of VN lexemes:

<table>
<thead>
<tr>
<th>nevent-vn-lex</th>
<th>agt-vn-lex</th>
<th>pat-vn-lex</th>
<th>aloc-vn-lex</th>
<th>ploc-vn-lex</th>
<th>cse-vn-lex</th>
<th>instr-vn-lex</th>
</tr>
</thead>
<tbody>
<tr>
<td>LECHE-VITRINE</td>
<td>GRATTE-PAPIER</td>
<td>GORE-MOUTON</td>
<td>JUREL-LOUP</td>
<td>GARDE-MEUBLE</td>
<td>COULE-SANG</td>
<td>TROTTE-BEBE</td>
</tr>
</tbody>
</table>

Table (9). Illustration of the different types of VN lexemes

Before looking at the detailed constraints on vn-lex, we must discuss the fact that the inheritance principled hierarchy allows us to adequately express the repartition between common and particular properties among the different VN lexemes we analyzed. And yet, the descriptive generalities hierarchically
ordered (as in fig. 5) fail to match the intuitive or desired picture according to which the most productive VN types should be ranked higher than the less productive ones. Indeed, considering the productivity of these lexemes, agent-\textit{vn-lex} should be the highest super type, or be the default type. However, that would lead to problems of descriptive congruency in the representation of other types (\textit{invent-\textit{vn-lex}}, \textit{loc-\textit{vn-lex}} and \textit{pat-\textit{vn-lex}}) as subtypes. A possible solution, to be explored in later work, would be to include a \textsc{productivity} feature with a variable of weight as value, that would integrate results obtained from a robust corpus study of VN lexeme productivity, based on Baayen’s (1992 and 2008) methods.

Constraints associated with the different \textit{vn-lexeme} subtypes integrate the general semantic rules proposed earlier. The constraints in (33) stipulate that a lexeme \textit{vn-lexeme} is a noun with a \textit{complex-nominal-relation} and two morphological daughters: a verbal base-lexeme and a nominal base-lexeme. The verbal base has a \textit{dynamic-rel} semantic relation and, by default, it has a \textit{patient-relation}, which means it minimally has a patient argument, and may possibly have an agent-patient relation. In addition, the value of its semantic argument \textsc{patient} is, by default, the same as the \textsc{index} of the nominal base-lexeme. Stem phonology is preceded by concatenation, in the standard way.

(33) \textit{vn-lexeme} =>

\[
\begin{array}{l}
\text{STEMS} [ \text{SLOT1 [3] } \oplus [4] ] \\
\text{SYNSEM} [ \text{CAT [ HEAD noun ] } ] \\
\text{CONT [ IND [1] ] } \\
\text{RESTR [ cplx-nom-rel ] } \\
\text{n-lex}\text{ M-DTRS } < v-lex \\
\text{STEMS [ SLOT1 [3] ] } \\
\text{SS \textsc{cont} [ NULC dynamic-rel [ PAT [2] ] ] } \\
\text{n-lex} [ \text{SLOT1 [4] ] } \\
\text{SS \textsc{cont} [ IND [2] ] } \\
\end{array}
\]

In (34), constraints on \textit{nominal-event-vn-lexeme} (\textit{nevent-vn-lex}) stipulate that the value of the \textit{austinian} \textsc{argument} of its \textit{relation} is the same as the \textsc{content} value of the verb base.

In (35), \textit{nominal-object-vn-lex} (\textit{obj-vn-lex}) has, by default\textsuperscript{8}, an \textit{austinian} \textsc{argument} in its set of \textsc{restriction}, whose value is the same as the \textsc{content} value of the verb base.

\textsuperscript{8} This default is overridden by \textit{cse-vn-lex} and \textit{instr-vn-lex}, which needs to introduce an additional agent argument here, \textit{via a causal-relation} or an \textit{instrumental-relation} respectively.
VN lexemes that express spatial localization (loc-vn-lex, in 40) have either an intransitive verb base (aloc-vn-lex, in 41) or a transitive one (ploc-vn-lex, in 42):

and doing so, embeds the austinian argument that corresponds to the content of the verb base.
VN lexemes that denote a patient are not productive, and not numerous. But when they correspond to a type of VN, the constraints needed are as below:

(43) \[ \text{pat-vn-lexeme} \Rightarrow \left[ \text{CONT} | \text{RESTR} \{ \begin{array}{c} \text{selectpat-rel} \\ \text{INST} \ [1] \\ \text{ARG} \ | \text{NUCL} \ [ \text{AGT} \ [2] ] \end{array} \} \right], [\text{IND} \ [2]] > \]

The figures below illustrate different lexical entries: in (44), LÈCHE-VITRINE is a type of vevent-vn-lex; in (45), GRATTE-PAPIER is an agent-vn-lex; in (46), HURLE-LOUP is a toponym, a type of aloc-vn-lex; in (47), COULE-SANG is a cse-vn-lex; and in (48), TROTTE-BÉBÉ is an instr-vn-lex.

(44) \[ \text{LÈCHE-VITRINE} \rightarrow \left[ \begin{array}{c} \text{STEMS} \ [\text{SLOT1} \ [3] \oplus [4]] \\ \text{SYNSEM} \ |	ext{CAT} \ | \text{HEAD noun} \\ \text{CONT} \ |	ext{RESTR} \{ \begin{array}{c} \text{lèche-vitrine-rel} \\ \text{INST} \ [1] \\ \text{ARG} \ | \text{NUCL} \ [\text{AGT} \ [2]] \end{array} \} \end{array} \right], [\text{IND} \ [2]] > \]
(45)  **GRATTE-PAPIER —>**

```
STEMS  [SLOT1 [3] ⊕ [4]]
SYNSEM [CAT | HEAD noun
         CONT [IND [1]
                RESTR { gratte-papier-rel
                           INST [1]
                           ARG [5] }]
M-DTRS < v-lex STEMS [SLOT1 [3]/gratte/]
                [SS | CONT [5] [NULC gratte-rel [ARGT [1], PAT [2]]]]
      n-lex STEMS [SLOT1 [4]/papier/]
                [SS | CONT [IND [2]]]
```

(45)  **HURLE-LOUP —>**

```
STEMS  [SLOT1 [3] ⊕ [4]]
SYNSEM [CAT | HEAD noun
         CONT [IND [1]
                RESTR { hurle-loup-rel
                           INST [1]
                           ARG [5] }]
M-DTRS < v-lex STEMS [SLOT1 [3]/hurle/]
                [SS | CONT [5] [NULC hurle-rel [ARGT [2], LOC [1]]]]
      n-lex STEMS [SLOT1 [4]/loup/]
                [SS | CONT [IND [2]]]
```

(45)  **COULE-SANG —>**

```
STEMS  [SLOT1 [3] ⊕ [4]]
SYNSEM [CAT | HEAD noun
         CONT [IND [1]
                RESTR { coule-sang-rel
                           INST [1]
                           ARG [5]
                           ARGT [1] }]
M-DTRS < v-lex STEMS [SLOT1 [3]/coule/]
                [SS | CONT [5] [NULC coule-rel [ARGT [1], PAT [2]]]]
      n-lex STEMS [SLOT1 [4]/sang/]
                [SS | CONT [IND [2]]]
```
4 Conclusion

We have presented here a formalized account of French Verb-Noun compounds, in line with the morphological analysis proposed in Villoing (2009). Our analysis integrates vn-lexeme types into the general lexeme typed-hierarchy, under a FORMATION dimension that allows the expression of a general classification among lexemes. We expect that the question of lexeme productivity may be solved by the integration of specific features into lexeme entries, as the result of a corpus study of VN productivity modeled on the methods of Baayen (2008). Moreover, we have shown that the fact HPSG allows semantics to be encoded as an independent resource is an advantage in capturing the general semantic patterns that are involved in the formation of several (de)verbal lexemes. In fact, there are other systematic lexical variations, which do not come under morphology, that also involve some of the general semantic types of relations we propose here. The very productive inchoative/causative verb pattern (TO INCREASE intrans/trans), for example, involves the causal-relation. Consequently, it is worth considering semantics as a lexical-sign dimension of classification in itself, as a way to encode in the hierarchy the fact that some semantic relations are lexically productive rules, available both for words and lexemes.

References


