Have Cutthroats Anything to Do with Tracheotomes?
Distinctive Properties of VN vs. NV Compounds in French

Fiammetta Namer
Université de Nancy 2 & UMR «ATILF»
Fiammetta.Namer@univ-nancy2.fr

Florence Villoing
Université de Paris 8 & UMR «Structures formelles du langage»
FlorenceVilloing@u-paris10.fr

Abstract

The aim of this paper is to compare word formation constraints that operate during the construction of compound lexemes in French; precisely, we contrast VN-structured lexemes (*ouvre-boîte$_N$:* open-tin $=$' *tin opener'*) with NV ones (*anthropophage$_A$:*°man-eat*(er) = 'cannibal'*). The question is whether both VN and NV are obtained through the same rule (and differ only with respect to the chosen components), or whether each type corresponds to a specific compounding process. Our study is grounded on large-scale corpora of VN compounds on one hand, and NV ones, on the other hand. In the first part of the paper we therefore motivate (1) our claim that bound roots, such as °*anthropo* and °*phage* in *anthropophage* can be assigned lexical categories (although they are no autonomous lexemes), and (2) our decision to assign precisely the verbal category to bound roots such as °*phage*. The second part of the paper is devoted to the NV versus VN properties investigation, according to three criteria: a) the distribution of the lexical category and the semantic values for VN and NV, b) the process type denoted by V, and c) the thematic roles plaid by N with respect to V. The comparison of VN and NV, according to these three criteria, leads us to suspect that NV and VN are indeed formed, in French, by two distinct morphological rules.

1. Introduction

1.1. Data

Within French lexical morphology, compounding processes form either nouns or adjectives. Either type can be obtained through a variety of rules, where two nouns (1a), two adjectives (1b), a verb and a noun (1c) are involved.

(1) a. [timbre$_N$-poste$_N$]$_N$:   stamp-post $=$ 'postage stamp'
        [homme$_N$-grenouille$_N$]$_N$:   man-frog $=$ 'frogman'

    b. [doux$_A$-amer$_A$]$_A$:   sweet-bitter $=$ 'bittersweet'

c. [coupe$_V$-papier$_N$]$_N$: cut-paper = 'paper knife'
[brise$_V$-coeurs$_N$]$_N$: break-heart= 'heartbreaker'
[perce$_V$-oreille$_N$]$_N$: pierce-ear= 'earwig'
(porte) [coupe$_V$-feu$_N$]$_A$: break-fire= 'firebreak (door)'

In parallel, the French lexicon includes lexical units such as (2), which often appear to be made up of two Combining Forms[1], originating from Latin or Greek. These combining forms appear to be two nouns (2a), an adjective and a noun (2b), two adjectives (2c) or a noun and a verb (2d)[2].

(2) a. [anthropo$_N$morphe$_N$]$_A$: human-shape = 'anthropomorphous'

b. [érythro$_A$cyte$_N$]$_N$: red-cell= 'erythrocyte'

c. [afro$_A$-cubain$_A$]$_A$: african-cuban= 'afro-cuban'

d. [anthropo$_N$phage$_V$]$_A$: human-eat= 'anthropophagous'
[mélo$_N$mane$_V$]$_N$: music-like= 'music lover'
[caverni$_N$cole$_V$]$_{A/N}$: cavern-live (ing)= 'cavernicole'
[infanti$_N$cide$_V$]$_{A/N}$: child-kill= 'infanticide'

Our aim is to examine compound nouns and adjectives formed according to the patterns illustrated in (1c) and (2d), in order to propose a contrastive analysis. This study defends the hypothesis that compounds obtained through Word Formation rules at play in (2) belong to the current French morphological system, despite the fact that some compounds are borrowed from ancient languages (eg. Latin and Greek), and that a large amount of them are common to several contemporary languages (mainly Romance, Germanic and Slavic ones, cf. Darmesteter(1877)). Their productivity is confirmed by the amount and variety of neologisms that are found in the media (*bibliophobe, tabacolâtre, Jocondovore*). Moreover, at least three arguments support the idea that neoclassical compounds do not come under Lexeme Formation Rules belonging to the ancient languages that provide their Combining Form components:

I. Compounds are often formed through the association of two Combining Forms of various origins, and thus not systematically coming from the same Latin or Greek language: French+Latin (*insecticide*), Greek+French (*macromolecule* 'macromolecule'), French+Greek (*bureaucrate* 'bureaucrat'), Latin+Greek (*planisphère* 'planisphere'), French Truncated Form+French, Latin or Greek (*cinéphile* 'cinephile', *anglophone, alcootest* 'drunkometer');

II. The emergence of an improper thematic vowel is often observed between the components: for instance, the thematic vowel "o", greek-specific, is also used at component boundaries originating from French (*franco-allemand* 'Germano-French', *anglo-saxon*) or from Latin, instead of the expected "i" vowel (*cérébro-*

---

[1] We borrow the term 'Combining Form' to Warren(1990), Fradin(2000) and Iacobini(1999). In the examples, Combining Forms are preceded by the '°' symbol ; on that matter, see section 2.3.
[2] A table, at section Appendix B, provides a literal definition of the mentioned neoclassical compound lexemes, according to the meaning of the Combining Forms they contain.

spinal 'cerebrospinal', *génito-urinaire* 'genitourinary') (Darmesteter(1894: 256));

III. Infringements of both Greek and Latin compounding rules are met in the formation of many neoclassical coumpounds. For instance:

i. Benveniste(1974:163-170) investigates the genesis of the noun *microbe*, in order to demonstrate that its constructed meaning ("little life", that is "microscopic organism") cannot be accessed from a Greek compounding pattern, which would lead to build the "short life" interpretation, which is incoherent from a scientific point of view.

ii. Darmesteter (1894:253) analyses the nouns *oxygène* 'oxygen' and *hydrogène* 'hydrogen', in order to show their semantic "ill" formation, given that, unlike in French, the Greek element corresponding to *gène* does not mean "which produces" (*oxygène*: "which produces acid", *hydrogène*: "which produces water") but instead, "which is produced".

More precisely, a Noun+Verb combining process emerges from examples (2d), similar to the French Verb+Noun compounding process[3], with two apparent differences, that is that I components come from the Greek or Latin lexicon , and II component ordering is reversed (in a neoclassical compound, governing (X) components are on the right, and governed (Y) ones on the left, resulting in the YX pattern, whereas 'ordinary' compounds are in the XY form). Our study is thus in line with traditional morphological analyses which identify a so-called "neoclassical" compounding results (eg. in 2d) in contrast with "ordinary" French component-based compounding ones (the traditional distinction was already current in the 19th century, cf. for instance Hatzfeld et al. (1890)). From this hypothesis, we now investigate differences and similarities between both VNs and NVs. Our objective is to identify and compare semantic and categorial constraints on these two types of compounds.

1.2. General issue

Assuming that lexical units in (2d) come under some NV compounding rule, our purpose is to determine whether VN and NV compounds are formed by two distinct compounding rules in French, or by one and the same rule. In other words:

- do VN and NV compounding rules correspond to a unique French lexeme formation process, with identical properties, except components origin and, consequently, components ordering[4]?
- do they correspond to two different lexeme formation processes in French, which do not form the same unit type, but depend on different semantic and categorial constraints ?

---

[3] Verb+Noun compounding exists in each Romance language : on that matter, see Rainer & Varela (1992), Scalise (1992), Villalava (1992)
[4] As many authors have remarked (see for instance Iacobini (2004), or Corbin (2005)), components ordering in « neoclassical » compounding is borrowed to ancient Greek and Latin.

1.3. Outline

To answer these questions, we have tried here to identify differences and similarities between VN and NV compounding patterns, especially through the contrastive study of both categorial and semantic constraints that are implied in these compounds' formation.

Our study is carried out from a corpus designed according to principles presented in section 2 ; in doing so, we also explain which theorical issues we had to face during the task of identifying components as Combining Forms, in NV compounds such as (2d). Section 3 is devoted to the comparison of VN and NV compound properties. We are led, in section 4, to the conclusion that the VN and NV compounds are formed by two distinct compounding rules in French, rather than a single one.

2. Data collection

2.1. VN corpus

This corpus contains 2-3000 VN compound lexemes. They have been gathered mainly from lexicographic sources: general language dictionaries, encyclopaedias. Dictionaries were chosen so that they are typical enough of both general and technical French lexica. Among them, there are: *Trésor de la Langue Française*, *Grand Robert de la Langue Française*, *Dictionnaire de la langue française* (Littré), *Dictionnaire Général de la Langue française* (Darmesteter and Hartzfeld).

Beside dictionaries, other sources have been used to collect VNs: Part-of-speech tagged corpora, where VNs could be extracted, and the Internet open corpus[5]. Also, newspapers, magazines and catalogues have also occasionally allowed us to enrich our VN corpus.

2.2. NV corpus

NV corpus belongs to the French lexicon, without being limited to a single speciality domain: limiting oneself to a specific domain is actually the risk to avoid when examining lexeme formation rules using combining forms.

Basically, two lexicographic sources have been used to build up this corpus: the machine-readable version of *Trésor de la Langue Française*, and the *Dictionnaire des structures du vocabulaire savant*. The latter, compiled by H. Cottez, groups and defines some 2,700 components used in learned words formation, be these components combining forms or learned affixes.

The first version of the NV corpus resulted from a semantic intuition about which of the combining forms should actually belong to the Noun and Verb categories. On the basis of this first corpus version, we have applied a series of (still tentative) principles in order to confirm component categories. We give an account of this analysis in section 2.3. In all, we have gathered about 500 NV compound adjectives and nouns,

---

[5] Many thanks to M. Plénat, who sent us so many examples.

which satisfy category criteria detailed below.

2.3.   *Identifying constituents in NV compounds*

Making up NV corpus raised important theoretical issues which are related to the identification of VN components. Several problems were posed:

i.   How to recognize a compound lexeme among all the morphologically complex lexemes formed out of combining forms. In other words, how to be sure that their constituting parts are base-lexemes, and not affixes[6]?

ii.   Consequently, how can 'verb' and 'noun' categories be identified for these base-lexemes, whose particular property is never autonomous in syntax?

2.3.1.   *Theoretical framework*

These questions make sens within the lexeme-based theoretical framework (Anderson(1992), Aronoff&Fudeman(2004), Booij(2005), Fradin(2003)). In this framework, basic units in lexeme (cf. Matthews(1991)) formation processes belong to one of the major categories: N, V, Adj, and are clearly distinguished from affixes, which are nothing but exponents of phonological rules. Within this approach affixes are not assigned a major category, unlike other theories (cf. Williams(1981), Lieber(1981), Lüdeling et al. (2002)). Our position is to extend lexeme-based morphology principles to bound roots, that is to assume that Greek and Latin Combining Forms can be N, V, or Adj categorized lexemes, in the same way as free roots are. This point of view fundamentally differs from those which admit the existence of Combining Forms, without assigning categories to them (cf. Rey-Debove(2004), and Cottez(1988: VII-XX), who claims "learned vocabulary forming elements" to be "signs", that is morphemes, but provides them with neither of base-lexeme nor affix status.
        On the other hand, we follow the theoretical line adopted (among others) by Corbin(1985), Warren(1990), Fradin(2000;2003) and Iacobini(2004), who make a distinction, among learned Combining Forms, between base-lexemes on one side and affixes on the other side. So, designing or NV corpus is based on the fact that Combining Forms have been identified as base-lexemes, in compounds words such as *anthropophage*.

2.3.2.   *Affix or Base-Lexeme (N, V, A)?*

• Are compound components in example (2) base-lexemes (N, V, A) or affixes?

This question is relevant because, in words such as *anthropophage*, Combining Forms properties (eg. *anthrop(o)* and *phage)* make them closely related to base-lexemes as

---

[6] Rather than « **lexeme** », we prefer to adopt the term of « **base-lexeme ».** According to authors, the former may include lexical units such as determiner, pronoun, preposition (cf. Aronoff & Fudeman (2004)), and these categories are irrelevant in Lexeme Formation.

well as to affixes.

Combining Forms look like affixes, because they are bound roots. They fulfill a structural position within lexical units, not within phrases. Combining Forms look like base-lexemes, because they have categorial and semantico-referential properties.

2.3.2.1.   *Semantic criterion*

This criterion relies on the dichotomy between "lexical sense" and "grammatical sense", frequently mentioned in the literature (cf. Corbin(2001) who applies it to French morphology, and Kleiber(1999), to lexical semantics)

A base-lexeme is an abstract lexical unit, with context-independent phonological form, syntactic category and meaning. Unlike affixes, which do not refer to entities but are given a semantic instruction, Combining Forms refer in a stable way to a referential category, a linguistically encoded concept: and so do base-lexemes. In other words, CFs carry a referential meaning ; they are able to identify concepts, but they are not able to name them within utterances (Corbin 85, Iacobini 99, Fradin 2000). However, there are serious drawbacks when attempting to use this criterion:

i.   identifying the component referential value depends on its translation and this translation lacks reliability. A big diversity is observed among dictionaries (and, sometimes, within a single dictionary) when looking up at a Combining Form translation: more especially, CF denoting processes (eg °*graph(o)*) are sometimes translated by a verb ("to write"), and sometimes by its corresponding deverbal noun ("writing"). And as far as the Robert Brio is concerned, for instance, both translations are given.

ii.   identifying the component referential value also depends on the morphologically constructed meaning of the compound they are part of, and on the semantic relation between the compound constituents. Now, semantic features provided by dictionaries are often unsufficient: either they are not precise enough, or they describe referents rather than meanings. The lexicographic definitions of *pleuronecte*, for instance (« type of flat fish such as sole, turbot, halibut…") and *bathyergue* 'bathyergus' («rodent mammal from Africa, whose behaviour is close to mole") prevent us from access to the morphologically constructed meaning, because they have to do only with the referent description. Actually, *pleuronecte* and *bathyergue* are named according to their behaviours: a *pleuronecte* swims (°*necte*$_V$) on one side (°*pleur(o)*$_N$); a bathyergue works (°*ergue*$_V$), that is, "digs", in depth (°*bathy* $_N$).

Moreover, this semantic criterion is hard to use when base-lexemes have undergone a grammaticalization process: in that case, they behave as affixes; and their semantic value vacillates between referential pole and instructional pole (cf. Amiot&Dal(2005) about °-*logue*).

2.3.2.2. *Categorial criterion*

This criterion relies on the base-lexeme property of belonging to one of the major part-of-speech categories: noun, verb, adjective, adverb. In other words, establishing that a CF belongs to one of these categories is sufficient to guarantee its base-lexeme status. However, as CFs never have syntactic realization, none of the usual identification syntactic tests (distribution in utterances, their syntactic behaviour and their inflectional marks) can operate in order to decide whether they are base-lexemes or not. To meet this specificity, other means are required. Four of them are presented here

- **Means 1**: CF can be assign a part-of-speech category according to its translation or interpretation in French. It simply gets the same category as its translation. But this solution is rather limited, due to translation uncertainty (see above).

- **Means 2**: A CF allomorph to a base-lexeme is clearly categorizable. But the allomorphy notion sometimes conceals suppletion phenomena, and relations between a base-lexeme and its hypothetical allomorph may lack transparency. We will not address this issue here (cf Haspelmath(2002:26-28) for a synthetic description of the issue, and Boyé(2006) for a presentation of the suppletion).

- **Means 3:** CFs which are selected as affixation bases are nothing but base-lexemes. From there, if an affixation process obeys constraints on its base such that only one base category is selected, then any CF used as a base for this process, automatically gets an unambiguous category (cf. Corbin (1985:62-64; 1987:182-sq)). However, only few affixes obeys such strong categorial constraints; so in general this technique cannot determine for certain the value of base categories.

- **Means 4**: CFs can also be assigned the lexeme status (and a category) according to the morphological construction in which they are found. In particular, the semantic relation type reveals its categorial counterpart. For example, in YX compounds where there is some typical argument-to-verb relation between Y and X, we can expect X to be a verb, and Y, to be a noun.

Through combination of the means just presented, we are able to identify verbal (X component) and nominal (Y component) Combining Forms in morphologically complex lexemes such as illustrated in (2d). Examples (3) and (4) give an account of the way we have proceeded to category identification.

(3)   *oculogyre*$_{ADJ}$[7]

    a.   °*ocul*(o)

    b.   °*gyre*

---

[7] From a lexicographic point of view, *oculogyric* qualifies muscles « that enables eyballs to turn ».

The CF °*ocul* occurs as base in the adjective *oculaire*; this allows to label it with the category "noun", via Means3. The reason is that suffix *-aire* exclusively forms adjectives from nouns: *pôle*$_{NOUN}$ → *polaire*$_{ADJ}$. Given that *oculaire* is formed by means of this process, from the base °*ocul*, the latter is not but a noun.

   Similarly, categorial constraints exerted by *-oire* and *-ion* French suffixes are used to provide (3b) with a grammatical category, thanks to Means3. In French, the adjective *giratoire* 'gyratory' (also found written *gyratoire*) is given the definition: "revolving around a point or axis"; and the noun *giration* 'gyration': "act or instance of turning". We will assume that the base of these lexemes, respectively formes by *-oire* and *-ion* suffixation, is the allomorphic variant °*gir(at)* of °*gyre*. Each of these suffixes preferentially selects preferentially verbal bases (*méritoire*$_A$ 'meritorious' < *mériter*$_V$; 'to deserve,merit' *agression*$_N$ < *agresser*$_V$ 'commit agression'), but not exclusively (*républicatoire*$_A$ 'republicatory'< *république*$_N$ 'republic'; *baladoire*$_A$ '°balladory'[8]< *balade*$_N$ 'ballad'; *zonation*$_N$ < *zone*$_N$). However, when the same sequence is used as base for both affixed lexemes, it is nothing but a verbal base (*revendicatoire*$_A$ 'revendicatory', *revendication*$_N$ < *revendiquer*$_V$ 'revendicate'). This is confirmed by a formal clue: When the base of a *-ion* suffixed noun (or that of a *-oire* suffixed adjective) ends with *-at*, then it is identified as the supine form of a Latin verb. So, as °*girat*/°*gyrat* results from a Latin verb form, it is indeed an allomorph of °*gire* or °*gyre*. And therefore, °*gyre* is a verb.

   So, Means3 allows us to identify both a nominal bound constituent (°*ocul*) and a verbal one (°*gyre*) in *oculogyre*$_A$. In other words, using Means3 may lead us to analyse *oculogyre* as a compound NV adjective.

(4)   *anthropophage*$_{ADJ}$

      a.   Means1

      b.   This hypothesis is confirmed by Means4

Means 1 makes us assume that °*anthrop* is a noun and that °*phage* is a verb; °*phage* has to be analysed as a verb, because its semantic relation with °*anthrop* is of predicate-argument type: *phage* denotes a process which applies on its proto-patient argument, realized by °*anthrop*.

   Finally, Means1 and Means4 cooperate to tell us that *anthropophage* is an NV compound adjective.

3.   Analysis: three criteria

Once the task of gathering NV and VN corpora has been carried out, the analysis core has been that of comparing "neoclassical" compounds NV, and French-components based compounds VN. Three criteria **(c1)** to **(c3)** served at that purpose:

(c1)   category and semantic values, for VNs and NVs ;

---

[8] *baladoire* qualifies « undecent » dances: the adjective originates from the musical genre « ballad ».

(c2) the type of process denoted by V verbs, for each compound type;

(c3) the semantic relation that holds between V verb and N noun

By crossing these three criteria, we ask questions such as: are VN and NV properties identical, in partial overlapping or in complementary distribution ? Are there lexical gaps, corresponding to unattested **(c1)** to **(c3)** combinations ?

3.1.   VN compound properties[9]

3.1.1.   Criterion C1: category and semantic types of VN compounds

- **Category**: VN compounding mainly builds nouns, and rarely adjectives. The latter are illustrated in (5):

(5)   (*papier*) *tue-mouches*$_A$: kill-flies (paper) = 'flypaper'
         (*porte*) *coupe-feu*$_A$:: break-fire = 'firebreak (door)'

- **Compound semantic type**

Most VN nouns denote **artefacts**:

(6)   *ouvre-boîte*$_N$: open-tin = 'tin opener'
         coupe-papier$_N$: cut-paper = 'paper knife'

Others refer to (human or animal) **animate entities**,

         -        either by their usual **function** or **occupation**

(7)   *garde-barrière*$_N$: guard-gate = 'level-crossing keeper',
         gratte-papier$_N$: scratch-paper = 'penpusher'

         -        or by some **characteristic**, **salient property** or **behaviour**

(8)   *trouble-fête*$_N$: disturb-party: 'killjoy'.
         (*garcon*) *rabat-joie*$_A$:°reduce-joy (boy) = 'spoilsport'
         perce-oreille$_N$: °pierce-ear ='earwig' *(insect)*

Finally, a few VN nouns describe either **locations**

(9)   *coupe-gorge*$_N$: cut-throat = 'cut-throat alley'
         garde-meuble$_N$: mind-furniture = 'store house'

         -        or **events**

---

[9] On this matter, see among others Villoing (2003) and Fradin (2005).

(10)   *lèche-vitrine*~N~: *lick-window* = 'window-shopping'
         (jouer à) saute-mouton~N~: leap-sheep = 'leapfrog' (play)

3.1.2.   Criterion C2: the type of process denoted by V

In a VN compound, the process type denoted by the verb is strongly constrained ; according to Vendler's (1967) terminology, and that of his successors (among others, Dowty(1979)), it can only be **dynamic,** that is compatible with linguistic contexts such as "être en train de V [be Ving] / se mettre à V [start to V]/ s'arrêter de V [stop Ving] ", cf. (11) (for a detailed presentation, see Villoing(2003)). On the other hand, V process type is very unlikely to be **stative** (12):

(11)   *coupe-papier* (°cut-paper 'paper knife')
         Jean est en train de couper le papier/ s'est mis à couper le papier /s'arrête
         de couper le papier.
         'Jean is cutting the paper/started to cut the paper/stops to cut the paper'.

(12)   *sait-latin (°know-latin)
         *Jean est en train de savoir le latin/s'est mis à savoir le latin/s'arrête de
         savoir le latin.
         *'Jean is knowing latin/started to know latin/stops to know latin'

3.1.3.   Criterion C3: V to N semantic relation

There are strong restrictions on the semantic relations between V and its participants; and consequently, on roles that N may play with respect to V. According to Dowty (1991) who refutes thematic role labelling, and proposes instead a Proto-Agent/Proto-Patient continuum, N corresponds quite exclusively to V **Proto-Patient**, within a VN compound. The reason is that N satisfies "change-of-state" and "affectedness" criteria (cf. also Foley&Van Valin (1984) and Jackendoff (1990) for a definition of these criteria).
         This analysis also applies for a small number of VNs, like in (13) that seem to be based on intransitive verbs and that for which N looks like the verbal Proto-Agent argument:

(13)   *trotte-bébé* (°toddle-along-baby = 'baby walker') ;
         pense-bête (°think-of-silly = 'reminder').

         However, the semantic of the verb is causative and N is involved as a Proto-patient

(14)   *trotte-bébé* 'causes/makes the baby to toddle along' ;
         pense-bête 'causes/makes the silly to think'.

Moreover, we can observe other (less frequent) semantic relations between V and N in VN compounds. For instance, we find rare cases (such as (15)) where N doesn't meet

Proto-patient criteria but rather Proto-agent ones.

(15)   *cuit-vapeur*: cook-steam (saucepan type)
         pousse-pied: push-foot (boat that we push with foot)

Besides, there are VNs such as (16), where N can be analysed neither as a Proto-patient
, nor as a Proto-agent of the verb. Rather, N seems to correspond to a temporal (16a) or
locative (16b) verb modifier.

(16)   a.   *réveille-matin*: wake-up-morning = 'alarm clock'
                grille-midi: scorch-midday (plant)
         b.   *croque-télé*: crunch-tv (tv tray)

3.2.   NV compound properties

3.2.1   Criterion C1: Category and semantic types of NV compounds

- **Category:** generally, NV compounding produces adjectives like (17) which
  function equally as nouns. (Kerleroux 1991, Kerleroux 1996, Fradin 1997)

(17)   *ventriloque*$_{ADJ}$ ('ventriloquist')

- **Semantic type**.

When only the nominal category is realized, NV may refer:

         -        mostly, to a **concrete object**:

(18)   *odontoclaste*$_N$ (cell which breaks (°claste$_V$) teeth (°odonto$_N$)) ;
         trachéotome$_N$ ('tracheotome': scalpel used to incise (°tome) the trachea
         (°trache(o)))

         -        to an **animate entity**,

(19)   *biographe*$_N$ ('biographer')
         notonecte$_N$ ('noctonect')

         -        sometimes, to an **event**:

(20)   *lipolyse*$_N$ ('lypolysis')
         infanticide$_N$ ('infanticide')

3.2.2.   Criterion C2: the type of process denoted by the verb

The process denoted by the verbal combining form may be

- either **dynamic**

(21) *°ambule_V* = 'to walk' (*noctambule_A*: 'late-night/night reveller')) *;*

- or **stative**.

Among **stative** processes, most verbs describe **spatial relations** (Talmy 2000, Vandeloise 1986).

(22) *°phore_{V,}* = 'to carry' *(mélanophore_N*: 'melanophore')*,
 °fère_V = 'to contain' (*carbonifère_A*: 'carboniferous').

Others are **feeling** or **emotion** verbs (according to Levin classification in (Levin 1993)):

(23) *°phile_V* = 'to like' *(russophile_{A.}* 'Russophil'*)*
 °mane_V = 'to like' *(mélomane_A* 'music lover'*)*
  °lâtre_V = 'to adore' *(wagnerolâtre_A* = who adores Wagner*)*,
 °phobe_V= 'to hate' (*anglophobe_A*: 'anglophobic')

Other **stative** processes are **perception** verbs

(24) *°op _V* = 'to see' (*nyctalope_A*: 'day-blind')

or verbs of **existence** (Levin & Rappaport 1995)

(25) *°cole_V* = 'to live' *(limicole_A*: 'limicolous'; *cavernicole_A)*
 °bie_V = 'to live' (*phyllobie_N* = (organism) living on leaves: 'green leaf weevil'),

We also found one verb of **light emission (**Levin & Rappaport 1995)

(26) *°luque* = 'to shine' (*noctiluque_A*: 'noctilucent')

3.2.3. Criterion C3: V to N semantic relation

In NV compounds, semantic relations between N and V belong to two distinct dimensions, according to the verbal process type.

- For **dynamic verbs**, Dowty's (1991) criteria are applied (as in the case of VN compounds). As a result, we have observed that here the noun can play different semantic roles wrt the verb:

  i.   N may fulfill the **Proto-agent** criterion:

(27) *psychogène_A*: 'psychogenic'

  ii.   (more frequently) it corresponds to a **Proto-patient**

(28)   *lipolyse*$_N$: 'lypolysis'

        iii.   It may also be the case that N **does not meet any** of the proto-role requirements

(29)   *ventriloque*$_A$ ('ventriloquist') ;
       noctambule$_A$ ('late-night/night reveller')
       héliotrope$_A$ ('heliotrope')

For **stative** verbs, Dowty role assignment criteria are not relevant. Rather, verb-to-noun relations can be expressed by means of Talmy Figure / Ground notions (Talmy(2000)[10]). So, when V denotes a **spatial relation**, then N refers:

    - mainly to the **Figure**: (ex 25),

(30)   *mélanophore*$_N$ ('melanophore')
       carbonifère$_A$ ('carboniferous')

    - or, sometimes, the **Ground**,

(31)   *vasiducte*$_N$ (= 'carrying vessel')

When V describes a **predicate expressing an emotion or a feeling**, it sets up a relationship between two participants: the experiencer and the stimulus. We propose then to identify these verbs as spatial relations: experiencers are interpreted as **grounds** and stimuli as **figures**.
    As far as we know, these verbal predicates only combine with **figure** denoting nouns: (ex 27)

(32)   *mélomane*$_A$ ('music lover')
       anglophobe$_A$ ('anglophobic')

4.   Results

By crossing criteria **c1** to **c3**, (cf. the comparative table in section Appendix A), we notice the following contrasts between VNs and NVs:
    NV compound nouns and adjectives cover a much broader spectrum than VN compounds, whatever the examined criterion:

- Categorial: NVs belong equally to ADJ or NOUN category (whereas VN compounds are almost exclusively nouns);

---

[10] Talmy (2000) investigates non-agentive relations, between participants of a so-called **spatial** or **localisation event**. Such events involve two main roles : the Figure and the Ground ; incidentally, other roles like Manner, Cause, or Path, may intervene. The **Figure** is identified as the moving or localized object, viewes with respect to another object (the reference objet : **Ground**).

- V process-type: V components in NV compounds describe either dynamic or stative processes, unlike in VN compounds.

- N to V semantic relation: in VN compounds, the N component is typically a proto-patient; but it fulfills a wider amount of semantic roles in NV compounds: proto-patient, proto-agent, or none, with respect to dynamic V components; and either figure or ground with respect to a stative V component which expresses a spatial relation.

In other words, NVs instanciate a wide range of combinations **c1-c3**, whereas VN compounds mainly correspond to those combinations where V is dynamic and N is proto-patient.

5. Conclusion

In some cases (when the verb is dynamic) VN and NV compounds seem to share similar categorial and semantic constraints, and this would lead us to conclude that only one compounding rule is at play and selects various types of components. However, VN and NV have clearly different behaviour when V is stative. In this case, it can be selected only as NV compounds governing component. Consequently, only the NV compounding rule can build spatial relations between N and V. 'Ordinary' morphology, by means of VN compounding, is unable to construct meaning equivalent to what NV compounding offers: syntactic patterns are required. There is, for instance, no other ways to express "melomane", but with a syntagmatic expression: "music mad".
      The results of this study lead us to think that two distinct compounding rules are at play in French. This conclusion raises two major questions:

- why is VN compounding such a constrained process?

- where do NV rules come from? Should we consider them as belonging to the French linguistic system, or are they borrowed from those languages where the components come from (ie Latin, Greek) ?

The latter question is part of a larger issue, in which we try to determine whether neoclassical compounding makes use of rules inherited from the very languages components originate, or not. Following comparativist authors (Darmesteter (1894)) (and unlike eg. Benveniste(1974)) the inheritance hypothesis seems to be the mainstream.
      Therefore, we tried to assess this assumption, in the specific case of NV compounds, where both N and V are of Greek origin. The underlying hypothesis is the following: if either nominal or adjectival NV compounds formation were known in ancient Greek, then we would have arguments in favour of borrowing; consequently we could provide the contemporary formation of this compounds type with an historical explanation.
      However, various studies focusing on compounding in ancient Greek

(Chantraine(1933), Smyth(1920), and recent studies by A. Ralli[11]) are divergent. Only Chantraine identifies an NV compounding pattern in ancient Greek. Conversely, Smyth and Ralli consider (for different reasons) that this formation did not exist in ancient Greek. This divergence in analysis show how weak is an hypothesis that assumes neoclassical compounds to be formed by means of borrowings of classical language rules.

But do Latin and Greek have no influence at all on modern compound formation? Works devoted to neoclassical formation in French all tend towards an hybrid answer, be they in favour of the above-mentioned inheritance hypothesis or not. This is shown, for instance, by the way A. Darmesteter (1894:252-256) analyses Greek-component based neoclassical compounds. A. Darmesteter show how these compounds come from some (often improper) re-analysis of ancient Greek to French, in a section where he laments the massive intrusion of new neoclassical compounds in the 19th century French lexicon, compounds that are often disrespectful to the rules of the langages they originate from. For instance, the creation of a neoclassical compound in 19th century French, such as *adénographie*$_N$ (*adenography* = 'written or drawn record of the glands') would go through the prior invention of the equivalent word the French compound would have in ancient Greek, if this equivalent existed (e.g. °αδενογραφια = *adénographia*), from components (included verbs) which were attested in ancient Greek (e.g. αδην$_N$ = *aden-* (gland), γραφω$_V$ = *-grapho* (to write)), and in compliance (real or presumed) with ancient Greek morphological rules.

The impact of Greek (as well as that of Latin, which functioned as intermediary) is thus unquestionable, but the role played by the French linguistic system is equally crucial. We can see, indeed, that neoclassical compounds formation rules have been elaborated by French speakers having an indeniable knowledge of ancient languages. This assumption that Greek has been re-analysed in French has been shared by various linguists all the way through the 20th century. It is found, for instance, in Benveniste(1974:170) and later, in Bouffartigue et Delrieu (1996). In further research, we still have to determine which has been the influence of the VN compounding pattern, which was already very productive at the end of the 18th century, upon the formation (or the reanalysis) of French NV neoclassical compounds.

6. References

Amiot, D. & Dal, G., 2005, « Integrating Combining Forms into a Lexeme-based Morphology», in *Fifth Mediterranean Morphology Meeting*, 15-18 sept, Fréjus

Anderson, S.R., 1992, *A-morphous Morphology*, Cambridge: Cambridge University Press.

Aronoff, M., & Fudeman, K., 2005, *What is morphology?*, Blackwell Publishing, Oxford.

Benveniste, E., 1974, "Formes nouvelles de la composition nominale", *Problèmes de linguistique générale II*, Paris, Gallimard, p. 163-176.

Booij, G., 2005, *The Grammar of Words: An Introduction to Linguistic Morphology*. Oxford, Oxford University Press.

Bouffartigue, J. et Delrieu A.-M., 1996, *Les racines grecques*. Paris, Belin.

Boyé, G., (2006), "Suppletion", *Encyclopedia of Language and Linguistics, 2nd Edition*, K. Brown, London/Amsterdam, Elsevier, p. 297.

Chantraine, P., 1933 (1979), *La formation des noms en grec ancien*. Paris, Champion

---

[11] Personal communication

(Klincksieck).

Corbin, D., 1985, „Les bases non-autonomes en français ou comment intégrer l'exception dans le modèle lexical", *Langue française* n°66, Paris, Larousse, pp. 54-76.

Corbin, D., 1987, *Morphologie dérivationnelle et structuration du lexique.* Lille, Presses Universitaires de Lille.

Corbin, D., 2001, "Préfixes et suffixes: du sens aux catégories", *Faits de Langue*, Paris, Ophrys, 15, pp. 41-69.

Corbin, D., 2005, "French (Indo European: Romance) " in Booij, Lehmann & Mugdan (eds) Morphology An International Handbook on Inflection and Word Formation, Berlin/New York, Walter de Gruyter, vol1, art. 121.

Cottez, H., 1988, *Dictionnaire des Structures du vocabulaire savant. Eléments et modèles de Formation*, 4ème édition. Paris, Dictionnaires Le Robert.

Darmesteter, A. 1894. *Traité de la Formation des mots composés dans la langue française, (1ère édition: 1875).* Paris: Honoré Champion.

Darmesteter, A., 1877, *De la création des mots nouveaux dans la langue française et des lois qui la régissent.* Paris, Wieveg.

Dowty, D., 1979, *Word Meaning and Montague Grammar.* Synthese Language Library, n°7, Dordrecht: Reidel.

Dowty, D., 1991, « Thematic Proto-roles and Argument Selection ». *Language*, 67(3), pp. 547-619.

Fradin, B. 1997, "Esquisse d'une sémantique de la préfixation en anti-." *Recherches Linguistique de Vincennes*, 26: pp. 87-112.

Fradin, B., 2000, "Combining forms, blends and related phenomena", in Dolsescha Ursula, Thornton Anna (eds). *Extragrammatical and Marginal Morphology.* Münich, Lincom Europa. pp 11-59.

Fradin, B., 2003, *Nouvelles approches en morphologie*, Paris, Puf.

Fradin, B. 2005,. « On a semantically grounded difference between derivation and compounding". In *Morphology and its Demarcations*, Dressler W. U., D. Katovsky & F. Rainer (eds) Amsterdam / Philadelphia: John Benjamins.

Foley, W. A., and Van Valin, R. D., 1984.*Functional syntax and universal grammar.* Cambridge: Cambridge University Press.

Haspelmath, M., 2002, *Understanding Morphology*. London, Arnold.

Hatzfeld, A., Darmesteter, A., et al, 1890, *Traité de la formation de la langue*, Livre premier, Paris, Delagrave.

Iacobini, C., 1999, "Distinguishing derivational prefixes from initial combining forms", in *Proceedings of First Mediterranean Morphology Meeting*, Mytilene (Grecia), 19-21 septembre 1997, G. Booij, A. Ralli, S. Scalise (eds.), pp. 132-140.

Iacobini, C., 2004, "Composizione con elementi neoclassici", in M.Grossmann / F. Rainer (a c. di) *La formazione delle parole in italiano*, Tübingen, Niemeyer.

Jackendoff, R., 1990, *Semantic Structures*. Cambridge Mass., MIT Press.

Kerleroux, F, 1991, "Les formes nues sont-elles simples ?", *Linx 25*, p. 21-44.

Kerleroux, F. 1996, *La coupure invisible: études de syntaxe et de morphologie*. Villeneuve d'Ascq, Presses Universitaires du Septentrion.

Kleiber, G., 1999, *Problèmes de sémantique - la polysémie en questions*. Villeneuve d'Ascq, Presses Universitaires du Septentrion.

Levin, B. 1993. *English Verbs Classes and Alternations*. Chicago, IL: University of Chicago Press.

Levin, B. & Rappaport H.M., 1995. *Unaccusativity*. Cambridge, MA: MIT Press.

Lieber, R., 1981. *On the organization of the lexicon.* Doctoral dissertation, MIT.

Lüdeling, A., T. Schmidt, et al., 2002, "Neoclassical word formation in German." *Yearbook of Morphology 2001*,pp. 253-283.

Matthews, P. H. 1991, *Morphology, An introduction to the theory of word-structure*, Cambridge University Press, Cambridge.

Rainer, F. & Varela, S., 1992, « Compounding in Spanish », *Rivista di linguistica* 4 (I), Pisa.

Rey-Lebove, J., 2004, *Le Robert brio.* Paris, Dictionnaires le Robert.

Scalise, S., 1992, « Compounding in Italian », *Rivista di linguistica* 4 (I), Pisa.

Smyth, H. W., 1920, *A Greek Grammar for Colleges* (libre accès à l'URL: http://www.ccel.org/s/smyth/grammar/html/toc.htm). New York, American Book Company.

Talmy, L. 2000, *Toward a cognitive semantics (2 volumes).* Cambridge, MA, Massachusetts Institute of Technology.

Vandeloise, C. 1986, *L'espace en français: sémantique des prépositions spatiales*, Les Editions du Seuil, Paris.

Vendler, Z., 1967, "Verbs and Times", *Linguistics and Philosophy*, Z. Vendler, Ithaca, Cornell University Press, pp. 97-121.

Villalava, A., 1992, « Compounding in Portuguese», *Rivista di linguistica* 4 (I), Pisa.

Villoing, F., 2002, "*Les mots composés [VN]N/A du français: réflexions épistémologiques et propositions d'analyse.*" Nanterre, Université de Paris 10.

Villoing, F., 2003, « Les mots composés VN du français: arguments en faveur d'une construction morphologique », *Cahiers de Grammaire 28*, « Morphologie et Lexique », pp. 183-196.

Warren, B., 1990. "The importance of combining forms" in Dressler W. U. et al. (eds.) 1990: 111-132.

Williams, E., 1981, "On the Notions `Lexically Related' and 'Head of a Word'." *Linguistic Inquiry*, 12,(2), pp. 245-274.

Dictionaries

Gaffiot, F., 2000, Le Grand Gaffiot - Dictionnaire Latin-Français. Paris, Hachette.

Bailly, A., 2000, Dictionnaire Grec-Français, Le Grand Bailly. Paris, Hachette Education.

Dictionnaire historique de la langue française, 1993, Dictionnaires le Robert, Paris.

Cottez, Henri, 1988, Dictionnaire des structures du vocabulaire Savant, Eléments et modèles de formation, Les usuels de Robert, Paris, 4e édition.

Le Grand Robert Electronique, 1994, CD-ROM.

Le Robert Brio, Analyse des mots et régularités du lexique, 2004, Dictionnaires Le Robert-VUEF, Paris.

Trésor de la Langue Française, Dictionnaire de la langue du XIXe et du XXe siècle, 16 volumes, 1971-1994, CNRS, Gallimard, Paris. (TLFi: http://atilf.atilf.fr/tlf.htm et CDROM: CNRS édition, 2004.)

Appendix A

	Noms				Adjectives
	VN/**NV** concrete +anim	VN/**NV** concrete -anim	VN/**NV** event	VN/**NV** location	
Dynamic V + patient N	trouble-fête *(killjoy)* garde-barrière *(level crossing keeper)* gratte-papier *(penpusher)* perce-oreille (earwig) **biographe** *(biographer)* **anthropophage** *(anthropophagous)*	ouvre-boîte *(tin opener)* coupe-papier *(paper knife)* trotte-bébé *(baby walker)* pense-bête *(reminder)* **odontoclaste** *(odontoclast)* **trachéotome** *(tracheotome)*	lèche-vitrine *(window-shopping)* saute-mouton *(leapfrog)* **lipolyse** *(lipolysis)* **infanticide** *(infanticide)*	coupe-gorge *(cutthroat)* garde-meuble *(store house)*	tue-mouches *(kill-flie')* rabat-joie *(spoilport)* coupe-feu *(firebreak)* **anthropophage** *(anthropophagous)* **oculogyre** *(oculogyric)*
Dynamic V + agent N		cuit-vapeur *(cook-steam)* pousse-pied [type of boat]			**psychogène** *(psychogenic)*
Dynamic V + other participant N	**notonecte** *(notonect)* **noctambule** *(night reveller)* **ventriloque** *(ventriloquist)* **pleuronecte** **bathyergue** *(bathyergus)* **xylographe** *(xylographer)*	réveille-matin *(alarm clock)* **électrographe** *(electrographer)* *grille-midi* [plant name] *croque-télé* (tv tray) **héliotrope** *(heliotrope)*			**plantigrade** *(plantigrade)* **ventriloque** *(ventriloquist)* **héliotrope** *(heliotrope)* **noctambule** *(late-night, night-time reveller)*
Stative V Spatial relation + Ground N		**vasiducte** *('carrying vessel')*			
Stative V Spatial relation + Figure		**mélanophore** *(melanophore)*			**carbonifère** *(carboniferous)*

N					
Stative V Feeling + Figure N					**mélomane***(music-lover)* **wagnerolâtre** *(who adores Wagner)* **russophile** *(russophil)* **anglophobe** *(anglophobic)*
Stative V – other N	**cavernicole**				**nyctalope** *(day-blind)* **cavernicole limicole** *(limicolous)* **phyllobie** *(living-on-leaves)* **noctiluque** *(noctilucent)*

Appendix   B

anthropophage$_A$	'anthropophagous'	*Who °phage$_V$ = 'eat' °anthropo$_N$ = 'human beings'*
anglophobe$_A$	'anglophobic'	*Who °phobe$_V$ = ''hates' °anglo$_N$ = 'the British'*
biographe$_N$	'biographer'	*He who °graphe$_V$ = writes' bio$_N$ = 'life'*
bathyergue$_N$	'bathyergus'	*(Sort of mole who) °ergue$_V$ = 'works' in °bathy$_N$ = 'depth'*
carbonifère$_A$	'carboniferous'	*Which °fère$_V$ = ''carries' °carboni$_N$ = 'carbon'*
cavernicole$_A$	'cavernicole'	*Which °cole$_V$ = 'lives' in caverne$_N$ = 'caverns'*
électrographe$_N$	'electrographer'	*(Instrument that) °graphe$_V$ = 'writes' by means of °électro$_N$ = 'electricity'*
héliotrope$_A$	'heliotrope'	*Which °trope$_V$ = 'turns' towards the °hélio$_N$ = 'sun'*
infanticide$_N$	'infanticide'	*Action of °cide$_V$ = ''kill' a °infanti$_N$ = 'child'*
limicole$_A$	'limicolous'	*Which °cole$_V$ = 'lives' in °limi$_N$ = ''mud'*
lypolyse$_N$	'lypolysis'	*Action of °lyse$_V$ = 'destroy' °lypo$_N$ = 'fat'*
mélanophore$_N$	'melanophore'	*(Cell that) °phore$_V$ = 'carries' °mélano$_N$ = 'melanin'*
mélomane$_A$	'music lover'	*Who °mane$_V$ = 'likes' °mélo$_N$ = 'music'*
nyctalope$_A$	'day-blind'	*Who °ope$_V$ = 'sees' at °nyctal$_N$ = 'night'*
noctambule$_A$	'late-night, night reveller'	*Who °ambul$_V$ = 'walks' at °noct$_N$ = 'night'*
noctiluque$_A$	'noctilucent'	*Which °luque$_V$ = 'glows' at °nocti$_N$ = 'night'*

notonecte$_N$	'noctonect'	*(Insect that) °necte$_V$ = 'swims' on its °noto$_N$ = 'back'*
odontoclaste$_N$	'odontoclast'	*(Cell that) °claste$_V$ = 'breaks' °odonto$_N$ = 'teeth'*
oculogyre$_A$	'oculogyric'	*(Muscle that enables) °ocul$_N$ = 'eye(ball)' to °gyre$_V$ = 'turn'*
phyllobie$_A$	'green leaf weevil'	*(organism that) °bie$_V$ = 'lives' on °phyll$_N$ = 'leaves'*
plantigrade$_A$	'plantigrade'	*Which °grade$_V$ = 'walks' on its °planti$_N$ = 'soles'*
pleuronecte$_N$	'pleuronecte'	*(Fish that) °necte$_V$ = 'swims' on one °pleuro$_N$ = 'side'*
psychogène$_A$	'psychogenic'	*which is °gène$_V$ = 'generated' by °psycho$_N$ = 'mind'*
russophile$_A$	'Russophil'	*who °phile$_V$ = 'likes' °russo$_N$ = 'Russians'*
trachéotome$_N$	'tracheotome'	*(scalpel used to) °tome$_V$ = incise °trache(o)$_N$ = 'trachea'*
vasiducte$_N$	'carrying vessel'	*°vasi$_N$ = 'vessel' which °ducte$_V$ = 'carries'*
ventriloque$_A$	'ventriloquist'	*who °loque$_V$ = 'speaks' with his/her ventri$_i$= stomach*
wagnerolâtre$_A$	'who adores Wagner'	*who °lâtre$_V$ = 'adores' Wagner$_N$*
xylographe$_N$	'xylographer'	*(He who) °graphe$_V$ = 'writes' on °xylo$_N$ = 'wood'*