

CLASSIFICATION, HEADEDNESS AND PLURALIZATION: CORPUS EVIDENCE FROM FRENCH COMPOUNDS*

MARIA ROSENBERG

Department of French, Italian and Classical Languages
Stockholm University
SE-10691 Stockholm
Sweden
maria.rosenberg@fraitia.su.se

Abstract: This study examines the relation between headedness and pluralization in French compounds attested in a newspaper corpus. It leans on the classification of compounds proposed by Bisetto and Scalise (2005) that distinguishes between subordinate, attributive or coordinate, and additionally, endocentric or exocentric. However, as to what item constitutes a compound in French, it follows the much more restricted definition of Corbin (1992), *inter alia*. My result indicates that French compounds exhibit double marking, i.e., both on the head and externally, as to pluralization. In a larger perspective, this result can indicate a tendency towards external pluralization in French.

Keywords: corpus, compound, French, head, plural

1. Introduction

The aim is to empirically investigate headedness in relation to pluralization in different types of French compounds. The data is drawn from a corpus of newspapers. Here, it must be pointed out that the plural marker is almost never phonologically realized in French compounds. A similar study using spoken French as empirical evidence would thus yield very different results. Hence, this fact can be assumed to lead to instability as to where to put the plural marker in written French, and in

* The author is thankful to the anonymous reviewers and Georgette Dal for many valuable comments and suggestions.

addition, serves to motivate the objective of this study, namely to examine the position of the plural marker of compounds by the use of corpus evidence. Since the delimitation of which items are to be classified as compounds in French is complex, this question will also be discussed.

This study was partly incited by a remark in Giorgi–Longobardi (1991, 129) that the position of the head in Romance compounds, being to the left, clashes with two general tendencies in Indo-European languages. The first tendency implies that inflections such as gender and number should be marked on the head of a compound. In contrast, the other tendency implies that these kinds of inflections should be marked at the end position of a word. Thus, the bearing of these two tendencies upon the actual plural marking of French compounds and its relation to the position of the head will be examined.

2. Classification, definition and delimitation of a compound

Bisetto and Scalise (2005) propose a new classification of compounds, which, unlike previous accounts (e.g., Bloomfield 1933; Haspelmath 2002 or Booij 2005), has the advantage of being based on one consistent criterion, and aims at universal validity. It has been tested for French, among other languages. The criterion in question is the grammatical relation, not overtly expressed, holding between the two constituents of a compound. This relation can, as in syntax, be of three types, namely subordinate, attributive and coordinate.¹ Subordinate compounds involve a complement relation, e.g., *apron string* or *cut throat*. Attributive compounds contain either an adjective plus a noun: the adjective expresses a property as in a modifier relation (e.g., *blue cheese*), or two nouns, where the non-head expresses, often metaphorically, an attribute of the head (e.g., *snail mail*). The non-head can be a verb as well, such as in *play ground*. In coordinate compounds, the constituents are linked by the conjunction *and*. All three types can be endocentric or exocentric, i.e., with or without a head constituent (cf. section 3 below).² Besides, neoclassical

¹ Baroni et al. (2006) distinguish between four types: coordinative, attributive, argumental and grounding. Thus, they seem to have split up the subordinate type into two. Their classification will however not be further addressed in this paper.

² Bisetto and Scalise (2005) are aware of borderline cases: e.g., *green-eyed* seems to contain another relation in addition to the attributive which holds between *green*

compounds, such as *hydrology*, which contain “semi-words” according to Scalise (1986), belong to the subordinate class, and what one usually calls “phrasal compounds”, e.g., [*floor of a birdcage*] *taste*, should be classified as attributive. Bisetto and Scalise (2005) also remark that further divisions of subtypes of the three compound classes can be made by the use of other criteria, such as the semantic relation between the constituents, their origin, their categorial status, and their internal relation.

In this study, a compound is defined as being formed of two lexemes, thus being itself a complex lexeme; as to French, either a noun or an adjective according to Namer (2005, 133). As Amiot puts it:

“Generally, compounds are supposed to display the following characteristics: (1) they are formed by morphological rules, (2) these rules associate lexemes and (3) the compounds always denote classes of entities.” (2005, 190)

The term ‘lexeme’ is taken in the sense of e.g., Lyons (1963); Matthews (1972) or Aronoff (1992b). According to Aronoff (*op.cit.*, 13), a lexeme is “a (potential or actual) decontextualized vocabulary word” and member of one of the open grammatical categories, N, V or A³ (as opposed to the closed grammatical categories which contain prepositions, conjunctions, determiners, pronouns, etc.). Consequently, lexemes are underspecified for inflection. Furthermore, because some lexemes only exist potentially, they do not necessarily have to be listed in the lexicon (Aronoff 1994), and the other way around, all listed entities are not lexemes, as Corbin (1992, 50) points out. In contrast, words, then, are their fully inflected counterparts, ready to be used in syntax.

Having introduced the notion of lexeme, some types which are classified as compounds by Bisetto and Scalise (2005) are ruled out from this study. They argue that morphological research favours [NN] and [AN] nominal compounds, at the expense of other types, such as compound adjectives ([AA] *bittersweet*, and [NA] *girl crazy*), and compounds containing e.g., adverbs (It. *sottosopra* ‘upside down’) or prepositions (Fr.

and *eye*, namely a subordinate relation between the realized head *-ed* and *eye*. Hence, it remains to decide which one of them should be considered primary.

³ Note that Aronoff (*ibid.*, note 28) includes adverbs in the category of adjectives, something which is not evident. As to French, Dal (2007) shows that adverbs in *-ment* should be considered as adjectives. Abbreviations used in this paper: A = adjective; Adv = adverb; Fr. = French; It. = Italian; lit. = literally; N = noun; P = preposition; Pref = prefix; Pro = pronoun; V = verb.

sans papier lit. ‘without documents’).⁴ However, following the definition of a compound above and the work of, *inter alia*, Amiot (2004; 2005), who shows that two-part constructions of these types are either prefixed derivations or syntactic formations, they are excluded from the class of true compounds in French, hence from this paper.

As to the delimitation between compounds and lexicalized phrases in French, I will follow Corbin (1992, 51). She claims that only those complex sequences which cannot be generated otherwise than by morphological mechanisms (i.e., lexical rules of composition) are compounds (cf. (1) in the quotation from Amiot (2005) above). So, sequences which have syntactic structure, such as *boit-sans-soif* ‘drinks without thirst’ or *bas-bleu* ‘blue-stocking’, belong to the syntactic module. Corbin (1992, 36–41) argues convincingly that all of the syntactic, referential, semiotic or deviation criteria given to delimit compounds from other complex constructions fail.⁵ If my interpretation of Corbin (1992) is right, this will lead to the exclusion of constructions of the types [AN] and [NA], in addition to the types already excluded above. In my data then, four types remain, namely [AA], [NN], [VN] and [VV], which are indisputably to be considered as compounds in French, with the reservation that not all of the [AA] and [NN] constructions are compounds (cf. 5.4 below).⁶ In my opinion, Bisetto and Scalise (2005) are thus generous in their classification of compounds, especially when it comes to French. This is not to say that their generous approach, as well as that of Rainer–Varela (1992) concerning Spanish, could not, in future studies, be applied to French.

⁴ According to Amiot (2005, 185), Scalise (1992) proposes an analysis which distinguishes between derivations and compounds, which both contain prepositions, by using endocentricity vs. exocentricity as a criterion: derivations are endocentric and compounds are exocentric. By this means, Ital. *sottocommissione* ‘subcommittee’ is a prefixed endocentric derivation (*sottocommissione* is a *commissione*, hence right-headed), whereas *sottotetto* ‘attic’ is an exocentric compound, formed by the preposition *sotto* ‘under’ plus the noun, *tetto* ‘roof’ (a *sottotetto* is not a *tetto*).

⁵ For a presentation and discussion of various criteria to define and delimit compounds, see also Rainer–Varela (1992) who, regarding Spanish compounds, add the criterion of system adequacy.

⁶ See also Fradin (2003, 196, 199) for a restricted view of French NN compounds, or Noailly (1990, 65–93) who treats all French NN-constructions (including subordinate, attributive and coordinate compounds) as syntactic, i.e., they are not compounds.

3. Definition of the morphological head of a complex word

The morphological head of a complex word is often defined as the item which determines the grammatical category and/or the major part of its meaning, thus combining distributional and semantic criteria (see e.g., Carstairs-McCarthy 1992, 20).⁷ For example, the semantic criterion decides that the [NN]_N compound *salades-santé* ‘health salads’ is left-headed, whilst the [NN]_N compound *maîtres-mots* ‘master words’ is right-headed. The distributional criterion has no relevance here, since both constituents are nominal. In contrast, the [VN]_N compound *ouvre-boîte* ‘tin opener’ lacks a head, and is exocentric. Here too, the distributional criterion is not useful. Only the semantic criterion makes it able to decide that the noun constituent *boîte* is not the head. Thus, in this paper, the definition of the head of a compound will be based solely on semantic grounds. In this manner, I am following the definition of the head of a compound given by Haspelmath (2002, 87). Likewise, Rainer and Varela (1992, 122) define the head-constituent of a compound as being its hyperonym. This latter definition is, in my opinion, the same analysis that Lesselingue (2003) uses to classify one type of [NN]_N compounds in French (see section 5.1 below), although she adheres to constructional morphology, originating from Corbin (1987) (cf. section 2 above), in which the notion of a morphological head has no place (see e.g., Namer 2005, 20–3). In this particular respect, i.e., by accepting the notion of head, the theoretical position adopted in this study differs from constructional morphology. The definition of a head used by Bisetto–Scalise (2005) is not explicitly stated, but in my opinion, it seems to be semantic, even though they base their classification of compounds on a, as they say “consistent”, grammatical criterion.

Generally (e.g., Selkirk 1983 and Scalise 1986 among many), Romance compounds are assumed to be left-headed (Fr. *projet-pilote*), as opposed to Germanic compounds which are right-headed (*pilot project*). In French, so-called neoclassical compounds, such as *océanographe*, are traditionally considered to be right-headed (Namer 2005, 22). In this paper, neoclassical compounds are left aside, since, in my opinion, their status and the delimitation within this type as compounds or derivations is open to debate, but see Namer (*ibid.*) who thoroughly argues for their

⁷ I do not intend to enter into a debate here either about the complexity of the morphological head, or about the potential problem it poses within morphological theories (see e.g., Haspelmath 1992 or Beard 1998).

being compounds. Besides, note that if they were treated as compounds in this paper, they would be right-headed with final pluralization. By this, their exclusion from this paper affects my results in a negative way. In Table 1 (section 5 below), they are classified as [PrefN] and [PrefA].

The strictly lexicalistic approach of Williams (1981) and Di Sciullo–Williams (1987) does not accept interface between syntax and morphology. It claims that the head of a complex word is always to the right following the Right-hand Head Rule. Given the RHR, French lacks compounds. Instead, we are facing phrases, reanalyzed as syntactic words. These syntactic objects have thus no morphological form and are not governed by morphological principles. In contrast, given Lieber’s (1992) syntactic approach to word formation, there is no difference between a morphological head and a syntactic one: the complex head is one and the same, generated by syntax, and equals a X^0 (see also Josefsson 1997). Hence, when it comes to French compounds, Di Sciullo–Williams (1987) and Lieber (1992) reach the same conclusion, namely that they have syntactic structure. Consequently, these two approaches have no place in my study, since they contradict both the definition and the delimitation of a compound used in this study. Note however that the constructional morphology approach of Corbin (1992) is strictly modular, like Di Sciullo–Williams (1987), but differs from the latter by treating French compounds as morphological formations. The weak lexicalist model elaborated by Ackema–Neeleman (2004) tolerates interface between morphology and syntax, but nevertheless posits that morphology is autonomous from syntax. Ackema and Neeleman (*ibid.*, 30–1) make a distinction between the syntactic head and the morphological head: they are two qualitatively different types of complex X^0 s, and their position may vary between syntax and morphology.

4. Corpus and data

The data of compounds used in this study is drawn from a corpus of newspapers⁸ and consists of 583 French items. Apart from the advantage to access a great number of plural compounds in context, a newspaper

⁸ The corpus, compiled by Umeå University, Sweden, consists of two Belgian newspapers: *La Libre Belgique* (17 volumes from February 1986 to November 1987, containing 1 033 295 words) and *Le Soir* (12 volumes from April 1988, containing 1 118 891 words), as well as two French newspapers: *Le Monde* (14 volumes from

corpus can, by its need to compress the text, be assumed as a favourable environment for compounds. The method used for the selection of items has been to search for two words separated by a hyphen. This was largely motivated by the limits of the corpus, but does present a considerable advantage as well. The use of a hyphen can be considered as a strong indication of a compound in French. Furthermore, a French compound which does not contain a hyphen, but is written as a simple word, generally takes a final plural marker. This latter type would therefore be of little interest for this study.

Furthermore, the selected items have a unique plural form. They are either (i) preceded by a plural determiner, e.g., (*des*) *grille-viande* ‘meat grills’, or (ii) have at least one of their parts in a plural form, e.g., *chanteurs-comédiens* ‘singers-comedians’. Note that a compound represented by several identical plural forms is counted only once. However, a compound consisting of the same lexemes, but in which the two words differ as to plural form, counts as two. This variation can be due to masculine and feminine forms, such as *doux-amers* vs. *douces-amères*, both ‘bittersweet’, to misprints or instability as to pluralization, e.g., *machines-outils*⁹ vs. *machines-outil* ‘tool machines’. Finally, note that this study takes a solely synchronic perspective, therefore diachronic and etymological matters are left aside.

5. Results

5.1. Various types of plural linguistic units in the data

After manual filtering, 374, of the total of 583 items drawn from the corpus, were judged to be true French compounds. Their distribution is statistically significant according to the chi-square test ($p < 0.001$). All of the items have been distinguished according to the classification proposed by Bisetto–Scalise (2005), and thereafter, according to grammatical category (Table 1).

1945 to 1988, containing 556 192 words) and *Le Monde Économique* (5 volumes from 1983 to 1988, containing 994 528 words), hence, a total of 3 702 906 words.

⁹ This pluralization is the correct one according to the dictionary *Le Trésor de la Langue Française informatisé*, *TLFi* (see <http://atilf.atilf.fr/tlf.htm>).

Table 1

Classification of the plural linguistic items drawn from the corpus

COMPOUNDS				EXCLUDED TYPES	
		Type	Number	Type	Number
Subordinate	endocentric	[NN]	114	[Adv/PN]	24
				[AdvA]	6
	exocentric	[VN]	41	[Adv/PN]	12
Attributive	endocentric	[NN]	107	[AN]	55
				[NA]	17
	exocentric			[NA]	8
				[AN]	4
Coordinate		[NN]	82		2
	endocentric	[AA]	28		
		[VV]	2		
TOTAL			374		
EXCLUDED TYPES	endocentric			[PrefA]	35
				[PrefN]	30
	exocentric			[ProV]	12
				[NN]	12
				[AA]	2
				TOTAL	209
TOTAL					
					583

Thus, the endocentric and exocentric subordinate, attributive and coordinate compounds are of the [NN], [VN], [AA] and [VV] types, all included in the box in Table 1. Following the standpoints discussed in sections 2–3 above, sequences containing adverbs and prepositions (see (1)), syntactic formations (see (2)), prefixed derivations (see (3)), as well as some [AA] and [NN] sequences which are not coordinated (see section 5.4 below), are all excluded.

- (1) [AdvN] avant-coureurs ‘forerunners’
[AdvA] moins-values ‘less valuable’
[PN] sans-abri ‘without shelters’
- (2) [AN] faux-nez ‘false noses’
[NA] coffres-forts ‘strong boxes’
[ProV] on-dit ‘one says’
- (3) [PrefA] pluri-annuels ‘perennial’
[PrefN] micro-secondes ‘micro-seconds’

[NN] compounds dominate in my data, as shown in Table 1. They can be of three types, given the classification of Bisetto–Scalise (2005). Still, this view is not shared by everyone. According to Corbin (1992), followed by Lesselingue (2003), French [NN] compounds are underspecified as to semantic interpretation. Lesselingue (2003) deals with what she denotes $[N_1N_2]_{N_3}$ compounds in French as a continuum. At one end, the N_3 is a hyponym of the class denoted by N_1 , e.g., *poisson-chat* ‘cat-fish’; here included are thus both subordinates and attributives in the terminology of Bisetto–Scalise (2005). At the other end, there are hybrids, e.g., *porte-fenêtre* ‘window-door’, which correspond to endocentric coordinate compounds following Bisetto and Scalise. [NN] constructions, such as *la physique-chimie*, ‘physics-chemistry’ with two distinct referents, are syntactic (Lesselingue 2003) (cf. section 5.4 below). Rainer and Varela (1992, 119) claim that left-headed [NN] compounds in Spanish are best described by the semantic criterion of conceptual unity. They distinguish between two types of Spanish [NN] compounds: subordinate and coordinate. Therefore, similar to Lesselingue (2003), the attributives belong to the subordinates. For example, Rainer and Varela (1992, 126) give *ciudad dormitorio* ‘bedroom community’ and *problema clave* ‘key problem’ as examples of subordinate compounds; the last compound is considered attributive by Bisetto–Scalise (2005). Nevertheless, Rainer and Varela (*ibid.*) comment on the adjectivization process going on regarding *clave*.

One remark can thus be made here, namely that there might be a reason to question the distinction between subordinate and attributive compounds, or at least further clarify it. We will see below that these two [NN] types show no difference as to their pluralization patterns.

5.2. Subordinate compounds

The 155 subordinate compounds attested in my data belong to two subtypes, endocentric [NN] compounds and exocentric [VN] compounds.

5.2.1. Subordinate [NN] compounds

Let us first look at the most frequent type attested in the corpus, the subordinate [NN] compound:

Table 2

Headedness and pluralization in subordinate [NN] compounds

SUBORDINATE [NN] COMPOUNDS				
Endocentric	Left-headed	$[N_{pl}N_{inv}]$	31	94
		$[N_{pl}N_{pl}]$	63	
	(Right-headed)	$([N_{inv}N_{pl}])$	(14)	(20)
		$([N_{pl}N_{pl}])$	(6)	(20)
(Total)			(114)	

- (4) Left-headed (a) $[N_{pl}N_{inv}]$ chèques-restaurant ‘dinner cheques’
 (b) $[N_{pl}N_{pl}]$ emballages-cadeaux ‘present wrappings’
- (5) (Right-headed (a) $[N_{inv}N_{pl}]$ hit-parades
 (b) $[N_{pl}N_{pl}]$ services-jobs)

A closer look at the 20 right-headed compounds shows that they are all of Germanic origin or calques. Therefore, their status as French compounds is highly controversial, and they should in fact be excluded, hence the parentheses in Table 2 above. Nevertheless, an odd pluralization occurs in six cases, i.e., an internal plural marker on the non-head as in (5b), which could be seen as additional support for the tendency of double marking in French (cf. also 5.3 below).

As to the true subordinate [NN] compounds, Table 2 shows that they are all left-headed. The most striking result here is that double marking occurs in a majority of these compounds, i.e., 67%, as in (4b) or *sacs-poubelles* ‘bin bags’. Recall however that the fact that the pluralization is almost never phonologically realized in French compounds probably plays an important role here. Nevertheless, this result is in opposition with the claim of Rainer–Varela (1992, 136) that Italian and Spanish seem to expose the same relative order of compounding and inflection: subordinate left-headed [NN] compounds (that is, subordinate and attributive in this paper) have the plural on the head. They remark that Scalise (1986) explains this phenomenon by “head-operation” following Hoeksema (1985). Furthermore, Rainer and Varela (1992, 136) observe that in some Spanish [NN] compounds, which are not coordinate, the non-head has a plural marker. This last case could, according to them, be taken as appositional, in order to avoid compounds being formed by inflected words (see Scalise 1986). However, this last problem is ruled out in a lexeme-based approach, to which this study belongs.

Corbin (1992, 48) remarks that the fact that the first constituent in *timbres-poste* ‘stamp’ is a hyperonym of the compound can semantically explain the internal inflection, and to predict the connection between lexicon and inflection should be easier than to deal with *timbre-poste* in a syntactic way. In this paper, inflection, i.e., pluralization, is supposed to occur after compounding. This is however not to say that I accept a level-ordering model along the lines of e.g., Scalise (1986), which is shown to be problematic (see e.g., Booij 1987). Instead, this study departs from the head-marking strategy of Stump (2001).¹⁰ There are actually three patterns of inflection available: head-marking, external marking and double marking (see Stump *op.cit.*, 96–137 for further details). Indeed, as we will see in this paper, all three patterns are valid for compounds in French.

Following the model outlined by Stump (2001), French plural subordinate [NN] compounds can thus be considered to exhibit both head-marking and double-marking. However, according to my results here, double marking dominates.

5.2.2. Subordinate [VN] compounds

The [VN]¹¹ compounds often refers to human, animal or plant agents or instruments,¹² but also to places, actions and results (Rosenberg to appear). They are, with few exceptions, masculine, a fact that Surridge (1985) derives from their structure.

Table 3

Headedness and pluralization in subordinate [VN] compounds

SUBORDINATE [VN] COMPOUNDS		
Exocentric	[V _{inv} N _{pl}]	35
	[V _{inv} N _{inv}]	6
Total		41

¹⁰ Stump defines the head relation as follows: “*b*[ase] is the head of a morphological expression *d*[erivative or compound] if and only if *d* arises from *b* through the application of a category-preserving rule” (*op.cit.*, 100).

¹¹ The second element is a noun in all the attestations in my data, but it can occasionally be an adjective, e.g., *taille-douce* ‘engraving’, or a noun with an adverbial function, e.g., *traîne-nuit* ‘someone who hangs out at night’.

¹² Six of the [VN] compounds in the data are agentive. Lieber (1992) notes that French lacks an overt and productive instrumental suffix, such as the English *-er*. In Rosenberg (to appear), I show that the [VN] compound is predominantly instrumental in Modern French, although the *-eur* suffix also has an instrumental extension.

The pluralization is externally exhibited (see (6)) or not overtly exhibited (see (7)).

- (6) [V_{inv}N_{pl}] garde-fous ‘handrails’
- (7) [V_{inv}N_{inv}] (a) porte-parole ‘spokesmen’
(b) compte-tours ‘revolution counters’

In (7b), the second noun has an internal plural, even when the compound is singular.

My result shows that French [VN] compounds as to pluralization, if any, exhibit external marking.¹³ By this, [VN] compounds, in conformity with subordinate [NN] compounds, seem to follow the second tendency, i.e., final pluralization, remarked by Giorgi–Longobardi (1991). Besides, Rainer and Varela (1992, 130) mention that in Spanish [VN] compounds, the internal plural of the second constituent seems to extend even to cases where it results in semantic oddness, therefore, this final *-s* may end up being an empty linking element.

5.2.3. Internal structure of [VN] compounds

Given the approach adopted in this study, the [VN] compound is a morphological lexeme-based construction. It is assumed to involve a purely semantic relation between a predicate and its participant, as Villoing (2002) shows. Its first constituent corresponds to a verbal stem (i.e., Fr. *thème*) as proposed by Marouzeau (1952), Vogel–Napoli (1995) and Villoing (2002). Other analyses proposed to account for the internal structure of [VN] compounds are briefly discussed below (see also Rainer–Varela 1992, 127–130) who account for more or less the same analyses of Spanish [VN] compounds).

- (8) An agentive nominal null suffix is added to the verbal element, the compound is thus left-headed: [[V + \emptyset]_N + N]_N (Rohrer 1977; Clements 1992; Kampers-Manhe 2000).
- (9) An agentive nominal null suffix is added externally, the compound is thus right-headed: [[V + N]_{VP} + \emptyset]_N (Lieber 1992).

¹³ When comparing the pluralization in my data with the one prescribed by e.g., *TLFi* (cf. n. 9) for [VN] compounds, the tendency towards final plural is obvious. In five cases where *TLFi* gives the plural marker as optional, it is realized, and five compounds, which are not supposed to take a plural marker, yet realize it.

- (10) The [VN] compound is preceded by a subject null NP, which makes it left-headed and determines its form and gender (Contreras 1985; Núñez Cedeño 1992).
- (11) The [VN] compound is a syntactic phrase, reanalysed as a word: $N \rightarrow VP$ (Di Sciullo–Williams 1987; Zwanenburg 1992).
- (12) The first element is underspecified for categorial features (Barner–Bale 2002).
- (13) The first element is originally an imperative form (Darmesteter 1967).

(8)–(10) are, in my opinion, problematic, since they all stipulate the existence of something that does not exist.¹⁴ Furthermore, (8) makes it hard to explain why the left head does not take a plural marker, and why French generally lacks nominalizations of this type, **un allume*.¹⁵ (9) gives rise to right-headed French compounds, a rare phenomenon. (10) attributes an adjectival function to the [VN] compound, something that it seldom has. The syntactic approaches to the French [VN] type in (9)–(11) explain its internal structure as one between verb and direct object. Rainer and Varela (1992, 129) argue as to Spanish [VN] compounds that the noun constituent always satisfies the internal argument position of the verb constituent. However, in some French [VN] compounds, the assumed syntactic relation is one between a verb and an external argument/subject, such as *gobe-mouton* ‘poisoned plant or pill destined to kill animals’, lit. ‘gobble’ (in the stem sense) + ‘sheep’. This relation can thus not be explained by a syntactic approach. Moreover, at least as regards (11), the syntactic irregularity of the presumed syntactic phrase, i.e., the lack of determiner inside the VP, is not explained. (12) falls within Distributed Morphology (see e.g., Halle–Marantz 1993 or Marantz 1997) and has, I think, difficulty to explain the semantic relation between the two constituents. (13) might be relevant from a diachronic perspective, but has no place in a synchronic theory.

Semenza et al. (1997) show that Broca’s aphasics, characterised by their problem to produce action words, tend to omit the first element in Italian [VN] compounds. Thus, their study seems to indicate that this element is considered to have verbal properties, and that [VN] compounds

¹⁴ Extensive critique has been raised against null elements in morphology (see e.g., Pullum–Zwicky 1991 and Anderson 1992).

¹⁵ By this characteristic, and by the lack of determiner before the noun constituent, the [VN] qualify, in my opinion as “synthetic” in Bloomfield’s (1933) sense. However, Bisetto and Scalise (2005) claim that this label, used e.g., for a compound such as *taxi driver*, does not apply to Romance, hence neither universally.

are not listed in the lexicon. Hence, this interpretation of their results can be taken as further evidence for the approach adopted in this study. In conclusion, I will however quote Rainer and Varela: “To sum up, we have seen that the enigma of Romance v-n compounds is far from having been solved to everybody’s content” (1992, 130).

5.3. Attributive compounds

Let me begin by excluding the 17 right-headed compounds in Table 4, since they are all Germanic borrowings or loan translations. However, note the double marking of plural which occurs in four cases, as in (15b) below (cf. also 5.2.1 above).

Table 4

Headedness and pluralization in attributive [NN] compounds

ATTRIBUTIVE [NN] COMPOUNDS				
Endocentric	Left-headed	$[N_{pl}N_{inv}]$	15	90
		$[N_{pl}N_{pl}]$	75	
	(Right-headed)	$([N_{inv}N_{pl}])$	(13)	(17)
		$([N_{pl}N_{pl}])$	(4)	
(Total)			(107)	

- (14) Left-headed (a) $[N_{pl}N_{inv}]$ camions-suicide ‘suicide lorries’
 (b) $[N_{pl}N_{pl}]$ secteurs-clés ‘key sectors’

- (15) (Right-headed (a) $[N_{inv}N_{pl}]$ liberty-ships
 (b) $[N_{pl}N_{pl}]$ boys-scouts)

We see that the attributive [NN] compounds are all left-headed, and that 75 (83 %) of them exhibit double marking of plural, such as in (14b) or *livres-labyrinthes* ‘labyrinth books’. This result of my study, just like my result for the subordinate [NN] compounds, indicates that the pluralization of French compounds tends to follow simultaneously both of the conflicting tendencies residing in Indo-European languages. Yet, since double marking is more dominant than head-marking, the second tendency seems to override the first. Furthermore, this result contradicts the purely syntactic approach by Lieber (1992), all too powerful, that the plural of $[NN]_N$ compounds is essentially marked on the left head. Recall that Rainer and Varela claimed the same as to Italian and Spanish (see

5.2.1 above). However, it must be admitted that the empirical evidence of this study does not contradict the model of Ackema–Neeleman (2004), which accepts the autonomy of morphology, but nevertheless tolerates its interface with syntax. Their model can account for head-marking, external marking and double marking in French plural compounds, because it stipulates that morphological structure and syntactic one can block each other, and allows complex words to have internal constituent structure and specific principles to apply to morphological complex units.

To sum up, my result concerning the plural attributive [NN] compounds speaks in favour of the tendency of double marking in French.

5.4. Coordinate compounds

I follow here the claim by Bisetto–Scalise (2005) that coordinate endocentric compounds are two-headed. Exocentric coordinate compounds should, according to them, be of the additive type. According to Haspelmath (2002, 89), additive compounds, such as Korean *son-pal* ‘hand’ + ‘foot’, are rare in European languages. In my data, I have found no attestation thereof.¹⁶

Table 5

Headedness and pluralization in coordinate compounds

COORDINATE COMPOUNDS			
Endocentric	Two-headed	[N _{pl} N _{pl}]	79
		[N _{pl} N _{inv}]	1
		[N _{inv} N _{pl}]	1
		[N _{inv} N _{inv}]	1
		[A _{pl} A _{pl}]	28
		[V _{inv} V _{inv}]	2
Total			112

In Spanish coordinate [NN] compounds, both heads have plural markers; this regular pluralization constitutes an evidence of their two-headedness (Rainer–Varela 1992, 125, 136). As shown in Table 5, except three coordinate [NN] compounds, e.g., *horloger-bijoutiers* ‘watchmaker-jewellers’, which will not be further remarked upon, the coordinated nouns and adjectives in my data also show double marking, as expected.

¹⁶ Rainer and Varela (1992, 126–7) give one example of an exocentric [NN] compound in Spanish, *este-oeste* ‘east-west’, which denotes a cardinal point.

(16) [N_{pl}N_{pl}] boulangers-pâtisseries ‘bakers-confectioners’

(17) [A_{pl}A_{pl}] lourds-légers ‘light-heavy’

They focalise on two aspects of the referent, and none of their constituents restricts the reference of the other, such as *danseurs-chanteurs* ‘dancers-singers’ and the examples in (16) and (17). It is therefore notable that Rainer and Varela (1992, 125) apply their definition of a head, based on a hyponymic relation between the compound and its constituents, to coordinate compounds as well. On this point, Lesselingue (2003), who separates [NN] compounds in which the first part is a hyponym of the compound from [NN] compounds in which properties of both constituents together yield the meaning of the whole compound, seems more accurate. Note that both Selkirk (1983, 16–9) and Lieber (1992, 55) consider English [AA]_A compounds, such as *bittersweet*, as right-headed. According to Lieber (1992), French [AA] compounds should be left-headed, but this is not the position taken here.

The excluded [NN] constructions to which I alluded in section 5.1 above contain nouns which are not coreferential, but relate to a preceding noun, e.g., *des relations médecins-malades* ‘the relations of doctors and patients’. This is obviously not a case of compounding: there exists no noun which denotes persons being simultaneously doctors and patients. The hyphen is instead used to shorten the phrase, the preposition *entre* ‘between’ seems to be omitted in the example above. An excluded [AA] construction is *blancs-bruns* ‘white-brown’, which refers to eggs, being either white or brown. Moreover, Rainer and Varela (1992, 131) separate on semantic grounds a type of [AA] construction from coordinate [AA] compounds in Spanish. This type can, in my data, be exemplified by *grand-ducales*, which is an adjectivization of the [AN] construction *grand-duc* ‘grand duke’, and hence excluded from my study.

The two coordinated [VV] compounds attested in my data are invariable, such as:

(18) [V_{inv}V_{inv}] passe-passe ‘tricks’

However, Rainer and Varela (1992, 127) note that, in Spanish, this type takes a final plural, and sometimes, two plural markers.

To conclude, the coordinate compounds in French, naturally, exhibit double marking.

6. Conclusion

This study examines the relation between headedness and pluralization in French compounds by the use of corpus data. The pluralization of French compounds exhibits all three patterns possible within Stump's (2001) head-marking model, namely head-marking, external marking and double marking. The left-headed subordinate and attributive [NN] compounds exhibit to a large extent double marking, and to a lesser extent head-marking. Subordinate exocentric [VN] compounds exhibit external marking, or no overt marking at all. The latter is also the case for coordinate [VV] compounds. As a result of their two-headedness, double marking is the only possible pattern for coordinate [NN] and [AA] compounds. According to my results, double marking is by far the dominant pattern. Thus, as to the two tendencies pointed out by Giorgi and Longobardi (1991), this study seems to indicate that the pluralization of French compounds follow both, i.e., head-marking and external marking, but that the second tendency seems to override the first, given that French compounds are predominantly left-headed. With respect to the subordinate vs. attributive distinction, a much more fine-grained analysis of left-headed French [NN] compounds would be of interest for future studies.

References

- Ackema, Peter – Ad Neeleman 2004. Beyond morphology. Interface conditions on word formation. Oxford University Press, Oxford.
- Amiot, Dany 2004. Préfixes ou prépositions? Le cas de *sur(-)*, *sans(-)*, *contre(-)* et les autres. In: *Lexique* 16 : 67–83.
- Amiot, Dany 2005. Between compounding and derivation. Elements of word-formation corresponding to prepositions. In: Wolfgang U. Dressler – Dieter Kastovsky – Oskar E. Pfeiffer – Franz Rainer (eds): *Morphology and its demarcations*, 183–95. John Benjamins, Amsterdam & Philadelphia.
- Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge University Press, Cambridge.
- Aronoff, Mark (ed.) 1992a. *Morphology now*. State University of New York Press, Albany.
- Aronoff, Mark 1992b. Stems in Latin verbal morphology. In: Aronoff (1992a, 5–32).
- Aronoff, Mark 1994. *Morphology by itself. Stems and inflectional classes*. MIT Press, Cambridge MA.
- Barner, David – Alan Bale 2002. No nouns, no verbs: Psycholinguistic arguments in favor of lexical underspecification. In: *Lingua* 112 : 771–91.

- Baroni, Marco – Emiliano Guevara – Vito Pirrelli – Eros Zanchetta 2006. Corpus evidence and compound structure: The case of Italian NN compounds. In: *Quantitative Investigations in Theoretical Linguistics 2 (QITL-2)*. University of Osnabrück, Osnabrück. (<http://www.form.unitn.it/~baroni/publications/qitl2006/qitl-compounds-handout.pdf>)
- Beard, Robert 1998. Derivation. In: Andrew Spencer – Arnold M. Zwicky (eds): *The handbook of morphology*, 44–65. Blackwell, Oxford & Malden MA.
- Bisetto, Antonietta – Sergio Scalise 2005. The classification of compounds. In: *Lingue e linguaggio 2*: 319–30.
- Bloomfield, Leonard 1933. *Language*. The University of Chicago Press, Chicago.
- Booij, Geert 1987. Lexical phonology and the organization of the morphological component. In: Edmund Gussmann (ed.): *Rules and the lexicon*, 43–65. Catholic University of Lublin, Lublin.
- Booij, Geert 2005. *The grammar of words: An introduction to linguistic morphology*. Oxford University Press, Oxford.
- Carstairs-McCarthy, Andrew 1992. *Current morphology*. Routledge, London & New York.
- Clements, J. Clancy 1992. Lexical category hierarchy and ‘head of compound’ in Spanish. In: Laeuffer – Morgan (1992, 151–66). 151–66.
- Contreras, Heles 1985. Spanish exocentric compounds. In: Frank H. Nuessel (ed.): *Current issues in Hispanic phonology and morphology*, 14–27. Indiana University Linguistics Club, Bloomington.
- Corbin, Danielle 1987. *Morphologie dérivationnelle et structuration du lexique*. 2 vol. Niemeyer, Tübingen.
- Corbin, Danielle 1992. Hypothèses sur les frontières de la composition nominale. In: *Cahiers de grammaire 17*: 25–55.
- Dal, Georgette 2007. Les adverbes en *-ment* du français : flexion ou dérivation? In: Fabio Montermini – Nabil Hathout (eds): *Morphologie à Toulouse, Actes du colloque international de morphologie 4èmes Décembrettes*, 121–47. Lincom Europa, Munich.
- Darmesteter, Arsène 1967. *Traité de la formation des mots composés dans la langue française comparée aux autres langues romanes et au latin*. Librairie Honoré Champion, Paris.
- Di Sciullo, Anna Maria – Edwin Williams 1987. *On the definition of word*. MIT Press, Cambridge MA.
- Fradin, Bernard 2003. *Nouvelles approches en morphologie*. Presses Universitaires de France, Paris.
- Giorgi, Alessandra – Giuseppe Longobardi 1991. *The syntax of noun phrases: Configuration, parameters and empty categories*. Cambridge University Press, Cambridge.
- Halle, Morris – Alec Marantz 1993. Distributed Morphology and the Pieces of Inflection. In: Kenneth Hale – Samuel J. Keyser (eds): *The view from building 20. Essays in linguistics in honor of Sylvain Bromberger*, 111–176. MIT Press, Cambridge MA.
- Haspelmath, Martin 1992. Grammaticization theory and heads in morphology. In: Aronoff (1992a, 69–82).

- Haspelmath, Martin 2002. *Understanding morphology*. Arnold, London.
- Hoeksema, Jack 1985. *Categorial morphology*. Garland, New York.
- Josefsson, Gunlög 1997. On the principles of word formation in Swedish. Lund University Press, Lund.
- Kampers-Manhe, Brigitte 2000. La composition en français et en néerlandais. In: *Leuvense Bijdragen* 89: 157–72.
- Laeufer, Christiane – Terrell A. Morgan (eds) 1992. *Theoretical analyses in Romance linguistics*. John Benjamins, Amsterdam & Philadelphia.
- Lesselingue, Chrystèle 2003. Les noms composés [NN]N « holonymiques »: illustration de la spécificité sémantique des unités construites morphologiquement. In: Bernard Fradin – Georgette Dal – Nabil Hathout – Françoise Kerleroux – Marc Plénat – Michel Roché (eds): *Les unités morphologiques*, vol. 3. Silexicales, 100–7. CNRS and Université Lille 3, Villeneuve d'Ascq: & SILEX.
- Lieber, Rochelle 1992. *Deconstructing morphology: Word formation in syntactic theory*. Chicago University Press, Chicago.
- Lyons, John 1963. *Structural semantics: An analysis of part of the vocabulary of Plato*. Blackwell, Oxford.
- Marantz, Alec 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In: *University of Pennsylvania Working Papers in Linguistics* 4: 201–25.
- Marouzeau, Jules 1952. Composés à thème verbal. In: *Le français moderne* 20: 81–6.
- Matthews, Peter Hugoe 1972. *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge University Press, Cambridge.
- Namer, Fiammetta 2005. *La morphologie constructionnelle du français et les propriétés sémantiques du lexique: Traitement automatique et modélisation*. Mémoire présenté dans le cadre de l'habilitation à diriger des recherches. Université de Nancy 2, Nancy. (http://www.univ-nancy2.fr/pers/namer/Publis/HDR_Fiam.pdf)
- Noailly, Michèle 1990. *Le substantif épithète*. Presses Universitaires de France, Paris.
- Núñez Cedeño, Rafael A. 1992. Headship assignment in Spanish compounds. In: Laeufer – Morgan (1992, 131–49). 131–49.
- Pullum, Geoffrey K. – Arnold M. Zwicky 1991. A misconceived approach to morphology. In: *Proceedings of the West Coast Conference on Formal Linguistics* 10: 387–98.
- Rainer, Franz – Solelad Varela 1992. Compounding in Spanish. In: *Rivista di linguistica* 4: 117–42.
- Rohrer, Christian 1977. *Die Wortzusammensetzung im modernen Französisch*. Gunter Narr, Tübingen.
- Rosenberg, Maria to appear. The polysemy of agent nouns: Diachronic, synchronic and contrastive evidence from French and Swedish. In: *Sprachtypologie und Universalienforschung* .
- Scalise, Sergio 1986. *Generative morphology* (2nd ed.). Foris, Dordrecht.
- Scalise, Sergio 1992. Compounding in Italian. In: *Rivista di Linguistica* 4: 175–98.
- Selkirk, Elisabeth O. 1983. *The syntax of words*. MIT Press, Cambridge MA.

- Semenza, Carlo – Claudio Luzzatti – Simona Carabelli 1997. Morphological representation of compound nouns: A study on Italian aphasic patients. In: *Journal of Neurolinguistics* 10: 33–43.
- Stump, Gregory T. 2001. *Inflectional morphology A theory of paradigm structure*. Cambridge University Press, Cambridge.
- Surridge, Marie E. 1985. Le genre grammatical des composés en français. In: *La revue canadienne de linguistique* 30: 247–71.
- Villoing, Florence 2002. *Les mots composés [VN]_{N/A} du français: réflexions épistémologiques et propositions d'analyse*. Doctoral dissertation, Université Paris X-Nanterre.
- Vogel, Irene – Donna Jo Napoli 1995. The verbal component in Italian compounds. In: Jon Amastae – Grant Goodall – Mario Montalbetti – Marianne Phinney (eds): *Contemporary research in Romance linguistics: Papers from the 22nd Linguistic Symposium on Romance Languages El Paso/Cd. Juárez, February 1992*, 367–81. John Benjamins, Amsterdam & Philadelphia.
- Williams, Edwin 1981. On the notions 'Lexically Related' and 'Head of a Word'. In: *Linguistic Inquiry* 12: 245–74.
- Zwanenbun, Wiecher 1992. Morphological heads, French "compounding" and Germanic "prefixation". In: Laeufer – Morgan (1992, 167–79).167–79.