

# Study on Formation of French Compound Nouns

Xiaoqin Hu

Beijing Language and Culture University, China

**Abstract.** The formation of compound nouns is complex but plays an important role in language creativity. It is always an important issue in linguistic research. This article aims to present the combining rules of French compound nouns in order to better understanding their formation process. A discussion about several problematic issues in studies of compound nouns are formed firstly: the separation of morphological compounding from syntactical compounding in definition of compounds, the distinction between derivation and compounding and the delimitation of constituents of compound nouns. Then, it provides a systematic analysis of combining constraints with different types of compound nouns. The analysis considers different natures of constraints: phonological, categorical and semantic. It permits to encounter a series of rules in formation of compound nouns and have a clear view of the compound nouns formation process.

**Keywords:** French; compound nouns; morphological compounding; syntactical compounding.

## 1. Introduction

The growth of compound nouns is largely associated with language creativity. A great number of new words come from compound nouns. This makes it difficult to update the dictionary duly and include new compound nouns in dictionary exhaustively. Thus, if the mechanism of compounding can be well understood, it will be a great contribution to lexicography, language teaching, translation, and a range of NLP applications (including automatic acquisition of vocabulary, automatic analysis of syntax and semantics, natural language generation, etc.). Much work has been done for studying compound nouns from different aspects: their definition, formation, semantics ..., but all these descriptions are still far from being satisfied face to the rapid creation and various structures of compound nouns. The purpose of this article is to analyze the combining rules of French compound nouns, as well as their phonological, categorical and semantic constraints. Several problematic issues about compounds are also discussed, like separation of morphological compounding from syntactical compounding, distinction between derivation and compounding, and delimitation of compound nouns constituents. Overall, there exist two different points of view about definition of compound nouns: the first considers any lexicalized nominal unit composed of two or more terms as compound nouns; the second distinguishes morphological compounding and syntactical compounding. Fradin (2009) even gave a definition of compounds very restricted, which is only reserved to morphologically-formed compounds. The analysis in this article focuses on compounding process of French nouns. The morphological and syntactical compounding are distinguished, but both of them are considered as formation process of compound nouns. Morphological and syntactical compounding process are presented separately, but the combination is not restricted between them for forming new words. In this article, it is necessary to clarify the definition of compound nouns and delimitation of their constituents firstly. Then, the combining constraints for forming compound nouns are analyzed from the point of view of phonology, category and semantics. Finally, we summarize the presentation and lead to the conclusion.

## 2. Definition of Compound Nouns

### 2.1 Definition of Compounding

Compound nouns are formed by compounding in the fact that two or several originally distinct terms, but are frequently occurring together in the syntax within a sentence, are attached as one unit which is difficult to analyse (Saussure, 1916, p. 242). They are groups of words which, taken together, have unpredictable semantics (Gross et al., 1986). Focusing on the formation process, the morphological compounding is distinguished from the syntactical compounding. Matthews (1991:

82), Corbin (1992: 28), Booij (2002: 141), Haspelmath (2002: 85) and Fradin (2003: 195) defined the compounding as a process of combing two lexemes rather in specialized language. They consider that the compounding only belongs to the lexical phenomenon but not to the syntactical one. In this theoretical framework, complex units composed of non-lexemes (ex., par-dessus, en dehors de...), complex conjunctions (ex., de sorte que, bien que...), lexicalized phrases (ex., qu'en-dira-t-on, rendez-vous...), lexicalized syntactic constructions are not considered as compounds (Villoing, 2012).

However, Guilbert (1971: LXIX), Mathieu-Colas (1996), Gross (1996: 33, 49, 53), Apothéloz (2002: 18-19), Di Sciullo (2005), Riegel, Pellat and Rioul (2009 :912) consider any complex lexical unit composed of two or more terms as a compound word. They consider N+prep+N units (such as brosse à dents, pomme de terre, femme de ménage...in French) as compounds. Gross (1986) described compound nouns on the basis of lexicon-grammar: "They are usually built from simple words by means of grammatical rules which may involve grammatical words". Mathieu-Colas (1996) developed this research and proposed a typology of compound nouns in French. He defined 17 elementary classes: LOANWORDS, ONOMATOPOEIA, Compound nouns with PARTICLES, N+de+X, N+en+X, etc. and 8 complex classes: EXPANSIONS de A+N, EXPANSIONS de N+N, etc. and distinguished two types of formations: superposition and fixed syntagms. These studies are criticized by Fradin (2009) on the ground that they confuse compounding with idiomaticity.

"It has become progressively apparent that compounds are problematic in models of grammar with segregated morphology and syntax" (Arnaud, 2015). English 'phrasal' compounds show that morphology may take in syntactic constructions (Lieber 1992: 14). Montermini (2008) sees in Italian units the objects that include syntactic sequences, which leads to the construction of models where the separation is less strict. In the framework proposed by Di Sciullo (2005), morphology and syntax can exchange word sequences and French compounds are derived in the syntactic space before being transferred to the morphology. Van Goethem (2012) has argued that N-clé (mot-clé, chiffre-clé, élément-clé...) was a morphological compound, but it develops a specific sense and is reanalyzed as adjective in the inherited construction N+clé in the process of degrammaticalization defined by Norde (2009: 120), which may be considered as a syntactical construction. In the construction grammar, "morphology and syntax are on a continuum, which solves the disputes over their delineation" (Arnaud, 2015).

## **2.2 Types of Compound Nouns**

Morphological compounding and syntactical compounding are generally considered as two different formation process of words. The compound nouns can be divided into morphological ones and syntactical ones. The syntactical compound nouns are formed from the lexicalization of syntactical constructions. Corbin (1992), Fradin (2003) and Villoing (2009) argue that VN compounds do not show the properties expected of lexicalized syntactic phrases but are rather formed morphologically, since they don't typically involve functional words such as determiners, prepositions, and pronouns and they don't preserve in their structure the original SVO word order of the source sentence. For morphological compounding, three types are distinguished: native compounding, neoclassical compounding and concealed compounding (Dal&Amiot, 2008). Native compounds are composed of two lexemes of current French lexicon without linking element (such as wagon-fumeur, vert-pomme, homme-grenouille, etc.) and neoclassical compounds are formed along two bases of Greek or Latin origin, that are not syntactically auto in French, connected by a linking element (such as microcéphale, homicide, cyno-céphale, etc.) (Villoing, 2012). The compounds formed of at least one truncated French lexeme (like afro- for africain in afro-brésilien, -ciel for logiciel in ludiciel, etc.) are called concealed compounds by Fradin (2000: 55). They are also called pseudo-compounds by Anderson (1992: 299). The other constituent of concealed compounds can be native or neoclassical, but concealed compounds are often grouped into neoclassical or native compounds.

Morphological compound nouns could also be divided into endocentric nouns and exocentric nouns depending on the semantic relations between two components. "The compound includes a head which transmits its semantic and syntactic properties to the compound is endocentric" (Villoing,

2012). The head is the semantic governing constituent of endocentric compounds. As well-known, the internal order of constituents in native compounds is XY and the internal order of constituents in neoclassical compounds is YX, where X is the governing component. The exocentric compounds present the absence of a head and this is the frequent type of compounds (Scalise&Vogel, 2010). The native compounds of NV pattern are exocentric, such as *casse-tête*, *ouvre-boîte*, *porte-manteau*, etc.

On the basis of the grammatical relations between the two constituents, the morphological compounds are classified into the followings: subordinate compounds, attributive compounds and coordinate compounds (Scalise&Vogel, 2010). The subordinate compounds are those of which two constituents have a relation of “complementation”, which is particularly evident in NN compound nouns, for example, *pause café*, *bébé-éprouvette*, *wagon-fumeur* ...; the two constituents of attributive compounds have a grammatical relation of attribution, typically AN or NN structures in French, such as *bleu ciel*, *électro-aimant*, *poisson-chat*, *homme-grenouille*, etc. The relation between two constituents of coordinate compounds is the relation of coordination, which often concerns native compounds of NN: *sapeur-pompier*, *horloger-bijoutier*, *histoire-géographie*...

### **2.3 Compounding and Derivation**

Both compounding and derivation are considered to belong to word formation, while the delimitation of boundary between compounding and derivation is problematic in defining the compound nouns. The definition along which the compound nouns are composed of free forms and the derivations are composed of bound forms raises many problems: the first problem concerns bound forms in neoclassical compounds and the second concerns free forms in the process of grammaticalization (Villoing, 2012). For example, the neoclassical compound noun in French *ludo-thèque* is composed of two bound elements which are not syntactically autonomous in French, but the neoclassical elements are free to combine with each other; if they are considered as affixes, the constructions like prefix+suffix without a lexical base defy the most fundamental definition of derivation. In the process of grammaticalization, some French lexemes *petit*, *grand*, *beau*... have been grammaticalized in certain lexical units like *grand-parents*, *petit-fils*, *beau-père*... They are considered as prefixes creating family terms. Many works have tackled this well-known issue and most of the studies point out that the boundaries between derivation and compounding are not distinct, since the border of the two processes can be crossed several times in the history of a langue (Mithun, 2010). Lesselingue and Villoing (2002) have shown that the complex lexemes like *aide-N* can be divided into two sets: VN compounds (*aide-mémoire*, *aide-nourrice*...) and prefixes (*aide-infirmière*, *aide-jardinier*...).

## **3. Constituents of Compound Nouns**

### **3.1 Word, Lexeme and Stem**

The lexeme is an abstract unit sub-specified for the flexion which corresponds to lexical units in linguistics, for example, *manger*, *mangeait* and *mangerai* are all the instances of lexeme MANGER (Fradin, 2003). The concrete words *manger*, *mangeait*, *mangerai*... are word forms of the lexeme MANGER. The word forms are the elements with syntactically relevant properties, while the lexemes are strictly morphological and have no syntactic realization (Villoing, 2012). The stem of a word is the word form minus its inflectional affixes, for example, *porte-* is the stem of *portais*, *portant*, *porta*, *porte*, *portons*, ... Each of the word forms of lexeme PORTER consists of a stem and a desinence (Booji, 2005).

Traditionally, it is considered that lexemes are the compounding units and native compounds are formed of two French lexemes without inflection, while the framework of lexeme-based morphology makes it clear to consider the verb in VN compound nouns as a stem of the lexeme (Corbin, 1992; Villoing, 2009; Fradin, 2009). However, the constituents of many compound nouns are words or even non-autonomous units. For example, in neoclassical compound nouns, the constituents are borrowed from Greek and Latin, which are non-autonomous elements and are not linked to any lexeme in French. They are called *archo-constituents* by Corbin and Paul (1999). Montermini (2010) argued

that archeo-constituents behave as lexemes, since they have a referential capacity and belong to the major categories (noun, verb and adjective). Some constituents of compound nouns are the truncated lexemes of French (such as, euro-, afro-, pétro-...). They are non-autonomous elements, which function as representation of the French lexeme, considered as fracto-constituents (Corbin and Paul, 1999). The fracto-constituents are treated as stems of lexemes, one of the forms of French lexemes, since they are linked to French lexemes (Montermini, 2010). Finally, some constituents reveal inflected forms of the lexeme, typically in VN compounds, such as oeufs in pique-oeufs, yeux in protège-yeux, mains in sèche-mains, etc. The word forms are also frequent in syntactical compounds, since the syntactical compound nouns are formed through the lexicalization of syntactical structures. Thus, most of the constituents of syntactical compounds have a syntactical realization, such as appareil à battre les collets, boîte aux lettres, boîte à déjeuner, trompe-l'oeil, etc.

A native compound can serve as the input for a new compounding process, like cure-dents in porte-[cure-dents]<sub>CN</sub>. This makes it interesting to ask: what does the notion “lexeme” used in definition of native compounds indicate exactly: simple lexeme or complex lexeme? If cure-dents is considered as native compound according to the definition of native compounding which combines two French lexemes without inflection, what status should be attributed to porte-cure-dents? If the notion “lexeme” used in definition of native compounds refers to either a simple lexeme or a complex lexeme, the compounding process of porte<sub>V</sub>-[cure-dents]<sub>CN</sub> corresponds to the definition of native compounds, in which cure-dents is a complex lexeme. In addition, neoclassical compounds could also serve as an input for a new compounding process, for example, poly[glycolyse]<sub>CN</sub>. It is the same for concealed compounds which are formed of at least one truncated French lexeme, that could be either simple or complex, for example, médico-[psychologue]<sub>N</sub>.

### 3.2 Syntagms and Complex Modifiers

Gross (1986) indicated that the variations of determiners and modifiers are accepted in compound nouns and that the compound modifiers (adjectives, noun complements and relative clauses) can be complex. This situation concerns syntactical compound nouns. They remain the syntactical constructions in their internal structure since they are formed from the lexicalization of syntactical constructions. However, what is the status of complex modifiers in syntactical compounding process? What are constituents of syntactical compound nouns?

The compound nouns of N1+PREP(à, de, sur, sans, en, pour, avec or par)+N2 are composed of a lexeme (N1) and a noun complement (N2) introduced by a preposition, such as in brosse à dents, pomme de terre, graveur sur pierre, chapeau sans bord, installateur en électroménager, crème pour les mains, appliques avec halogène, assemblage par boulons, etc. However, the element in the position of N2 is not limited to a simple noun, it could be a verb (such as machine à [laver]<sub>V</sub>, aiguiller à [brider]<sub>V</sub>, machine à [imprimer]<sub>V</sub>...), a noun phrase (such as boîte [aux lettres]<sub>NP</sub>, contrôleur [des risques financiers]<sub>NP</sub>, crème pour [les mains]<sub>NP</sub>...), a compound noun (such as, brosse à [tapis longs poils]<sub>CN</sub>, un mécanisme à [bouton poussoir]<sub>CN</sub>, bol à [bec verseur]<sub>CN</sub>...) or even a VO (verb+objet) construction (such as appareil à [battre les collets]<sub>VO</sub>...). Besides, two noun complements could be coordinated for modifying one head noun, for example, in conducteur de grues et d'engins de lavage, the two noun complements grues and engins de lavage are coordinated for modifying the head conducteur. Gross (1986) defined noun phrase by the basic form “DET+N+Modifier” and in Modifier, the N could be a simple noun, a compound noun or even another noun phrase. This is still evident in the structure of syntactical compounds, for example, in conducteur d'une machine à imprimer, the modifier d'une machine à imprimer includes a compound noun machine à imprimer. In addition, N1 in N1+PREP(à, de, sur, sans, en, pour, avec or par)+N2 have the same nature with N2, it could be a simple noun, a compound noun, a noun phrase or two coordinated nouns (noted as Cdt), for example, [directeur financier]<sub>NP</sub> du cabinet and [technico-commercial]<sub>CN</sub> en agrofournitures. [analyste et contrôleur]<sub>Cdt</sub> des risques financiers. It is the same situation for N in the syntactical compound nouns of the type N1+N2 (ampoule [basse consommation]<sub>NP</sub>), N+A ([pied plat]<sub>CN</sub> valgus statique), N+Pprés (thé dansant), N+Pp (colis piégé) and VO (trompe-[l'oeil]<sub>NP</sub>). It should also be pointed out that the

complex adjectives are also accepted as argued by Gross (1986), such as *technicien* [mécanique et électronique]<sub>Cdt</sub>, *appareil* [électro-acoustique]<sub>CA</sub>...

These analyses make us to ask the question: is it still relevant to consider words or lexemes (simple or compound) as constituents of syntactical compounds? Should the words in a noun phrase of a compound noun be considered as constituents of the compound noun or constituents of the noun phrase? For the compounds of N1+PREP+N2, N1+N2, N+Pprés, N+Pp and N+A, they could be treated as a noun phrase structure (Head+Modifier) defined by (Gross, 1986) without determiner, in which the head could also be a noun phrase itself. Generally, in N1+PREP+N2, N1 is the head, PREP+N2 is the modifier (a noun complement introduced by a preposition). In N1+N2, either N1 or N2 could be the head and the other component is the modifier which can be simple or complex. It is more evident for N+A, in which N is the head and A is the modifier. For compound nouns of the type VO, they belong to the verbal construction verb + objet, in which the object is a noun phrase in syntactic relationship with the verb. Thus, it is reasonable to consider syntagms as constituents of VO compound nouns.

## 4. Constraints of Combining

### 4.1 Phonological Constraints

Neoclassical compounds are defined as a combination of one base of Greek or Latin origin, which are syntactically non-autonomous in French, with a linking element or a liaison vowel /i/ or /o/ in certain phonological contexts. The vowel /i/ appears in Latinate components when the second components are Latinate constituents or fracto-constituents, for example, *homilatincidelatin* and *ludilatincielfracto*, while the vowel /o/ appears in Latinate components when the second components are Greek constituents or French lexemes, such as *ludolatinthèquegreek* and *insectolatinfongicidefr*. /o/ appears in Greek components (*anglogreekphonegreek*) or fracto-constituents (*afro-brésilien*) regardless of the nature of the second constituent. Either /i/ or /o/ could appear in French lexemes or proper nouns (*riziculture* and *sarkophobe*). Besides, /o/ is omitted before /i/, /a/ and /o/, such as *pyschiatre* (*pyscho-*, *-iatre*), *pédiatre* (*pédo-*, *-iatre*), *pyschanalyse* (*pycho-*, *analyse*), *hydroxyle* (*hydro-*, *oxyle*), etc. and /y/ appears in some Latinate constituents (*manucure*, *sinu-vertébral*, *natu-majorité*...). In some Greek components, /a/ appears, such as *bêta-lactamine*, *gamma-caméra*, *méga-complétude*...

However, it seems that this phonological liaison doesn't restrict the combination of two components of different nature (Latin, Greek, fracto-constituents or French lexemes): the neoclassical compounding combine archeo-constituents (*macro<sub>greek</sub>-céphale<sub>greek</sub>*, *ludo<sub>latin</sub>thèque<sub>greek</sub>*, *insecti<sub>latin</sub>cide<sub>latin</sub>*...), archeo-constituents and fracto-constituents (*ludi<sub>latin</sub>ciel<sub>fracto</sub>*, *euro<sub>fracto</sub>-phile<sub>greek</sub>*...), archeo- or fracto-constituents and French lexemes (*photo<sub>greek</sub>sensible<sub>fr</sub>*, *vélo<sub>fr</sub>drome<sub>greek</sub>*...) as well as French lexemes (*alcool<sub>fr</sub>-thérapie<sub>fr</sub>*, *droit<sub>fr</sub>-o-fasciste<sub>fr</sub>*...) (Villoing, 2012). However, elements with the liaison vowel /o/ are more productive, since they can appear before both Greek components and French lexemes. Thus, when a French compound noun (regardless of the nature of constituents) serve itself as an input for a new compounding process, the elements with the liaison vowel /o/ are often found, such as in *oto-rhino-laryngologiste*, *médico-psychosociologue*, *pharmaco-épidémiologie*...

### 4.2 Categorical Constraints

The native compound nouns mainly have three types of patterns: N+N (*appareil photo*, *timbre-poste*, *jeux vidéo*...), A+A (*clair-obscur*, *doux-amer*...) and V+N (*porte-manteau*, *abat-jour*, *allume-cigare*...). The patterns like V+V (*cache-cache*, *pousse-pousse*...), ADV+N (*malchance*, *malheur*...) and ADV+A/ADV (*pète-sec*, *gagne-petit*, *couche-tard*, *passe-partout*...) are no longer productive as indicated by Villoing (2012). For neoclassical compound nouns, there are N+N (*ludothèque*, *vélodrome*...), A+N (*micro-organisme*, *mégalithe*...) and N+V (*homicide*, *océanographe*...). As argued in section 3.1., N and A could be simple or complex, such as *arbren<sub>N</sub> porte-galet<sub>CN</sub>*, *medico<sub>A</sub>-psychologue<sub>CN</sub>*, etc., while in the pattern of neoclassical compounds N+V, N is rarely complex and

in A+N, A is frequently simple. Besides, a native compound noun can function as a constituent for forming a new native compound noun ([[maître-nageur]<sub>nativeCN</sub> sauveteur]<sub>nativeCN</sub>), but it rarely serves as an input for forming a neoclassical compound. Nevertheless, a neoclassical compound can serve as input for forming both a native and neoclassical compound noun ([médico-[psychologue]<sub>neoCN</sub>]<sub>neoCN</sub>, [appareil[électro-acoustique]<sub>neoCN</sub>]<sub>nativeCN</sub>...). However, when we compare the outputs of two compounding processes, it is found that there are both adjective outputs and noun outputs for neoclassical compounding by patterns NN, AN and NV ([cynoncépahlen]<sub>A</sub>, [longialignen]<sub>A</sub>, [anglonphilev]<sub>A/N</sub>...), while the outputs of native compounding by NN and VN are usually nouns. It is less certain for forming compound nouns with the patterns of neoclassical compounds.

The main patterns of syntactical compound nouns are the followings : N1+PREP+N2/V/VO (fil de fériste, joie de vivre, appareil à battre les collets...), N1+N2 (graphiste multimédia, producteur télé, ampoule basse consommation...), N+A (cordon-bleu, table ronde, cuir-chevelu...), N+Pprés (cerf-volant, chat-huant, toit ouvrant...), N+Pp (lettre recommandée, banane flambée, compte-rendu...), VO (crève-la-faim, barre-la-route, trompe-l'œil...), V+PREP+N/A/Divers (touche-à-tout, pince sans rire, saute en bas...), V+et+V (va et vient, aller et retour, lève et baisse... ) and A+et+A (noir et blanc, chaud et froid, bleu et blanc...), in which N could be simple, complex or a noun phrase and A could also be complex. However, in A+et+A, A is rarely complex and V+PREP+N/A/Divers, V+et+V as well as A+et+A are less productive than the other patterns of syntactical compound nouns. The morphological compounds are allowed in syntactic constructions (année [luni-solaire]<sub>C</sub>, système [hypothético-déductif]<sub>C</sub>, [couteau [lame trapèze]<sub>C</sub>]<sub>C</sub> à [bouton poussoir]<sub>C</sub>...), while the syntactic constructions are not accepted in the definition of morphological compounding. It seems that morphological compounding is limited to the combination between different types of morphological compounds (native, neoclassical or concealed). Nevertheless, if a syntactical compound serves as input in a morphological compounding, for example, the syntactical compound filtre à café in [porte]<sub>V</sub>- [filtre à café]<sub>N</sub> which corresponds to native compounding VN, what status should be attributed to the obtained compound noun? Is it still a morphological compound noun? filtre à café is only a complex lexeme in porte-filtre à café, since it is a lexicalized syntactic construction. In this case, porte-filtre à café is still considered as a native compound noun. Overall, it doesn't exist a distinct combining restriction between the syntactical compounding and morphological compounding. On the contrary, their exchange in combining makes it more productive for word formation.

### 4.3 Semantic Constraints

#### 4.3.1. Coordinated Compound Nouns

The compound noun N1+N2 can be coordinated such as horloger-bijoutier, linguiste informaticien, sapeur-pompier, rhino-pharynx, technico-commercial, ... They denote countries (Bosnie-Herzégovine), fields of study (histoire-géographie) or profession (technico-commercial). These compound nouns are often taken in another coordinated NN for forming a new compound noun, such as maître-nageur in maître-nageur sauveteur, sapeur-pompier in infirmier sapeur-pompier, rhinolaryngologie in oto-rhino-laryngologie, hydrogène in hydrogénocarbonate..., while this type of pattern is more productive for forming native compound nouns than neoclassical or concealed ones.

In native compound nouns of A1+A2, the two adjectives are often antonyms (clair-obscur, doux-amer, chaud-froid...). They belong to the vocabulary for describing one same property, such as gris-bleu (both gris and bleu belong to the vocabulary of color), sourd-muet (both sourd and muet are the vocabulary for describing two sense organs) ... If an object or an alive has different properties and each property (e.g. age, color, size...) have different values (e.g., age|[elder, young...], color|[blue, red, yellow...], size|[big, small, medium...]), the adjectives in A1+A2 refer to different values of this property. A1+A2 can be used as an input for forming a new compound noun of type AA and A in A1+A2 can be complex, but limited to the AA compound (sourd-muet-aveugle). For neoclassical compound nouns, most of the neoclassical compounds AA are adjectives, and like native compounds of AA, their constituents also refer to different values of a property. Nevertheless, the neoclassical compounds of AA rarely serve as an input for forming native compounds of AA and the native compounds of AA are also rarely used as constituents in neoclassical compounds of AA either. This

restriction is due to the semantic relation between two components of AA compounds and the origin of neoclassical compounds. AA compound include two adjectives which describe the two different values of a property, while neoclassical compounds belong to a scientific nomenclature and native compounds are mostly non-scientific. Thus, it hardly exists the possibility of combining between the neoclassical compounds and native compounds of type AA.

#### 4.3.2. Attributive Compound Nouns

In attributive compound nouns of N1+N2, the governed N generally specifies a part, a property or a function of the noun head, such as poisson-chat (chat indicates a species of fish), chou-fleur (fleur describes a property of cabbage) .... In native compound nouns, N1 is the governing noun and in a neoclassical compound noun, the order is reversed (hydrog<sub>N1</sub>carbonat<sub>N2</sub>). N in N1+N2 could be complex (arbren [porte-galet]<sub>CN</sub>), but the attributive compounds of N1+N2 are rarely used as constituents for forming native compounds. The combining ability of this type of compound nouns is very limited.

In attributive compound nouns of A+N, it is not N1 which specifies the property of N2 but A, for example, micro<sub>A</sub>organism<sub>N</sub> (micro indicates the scale of the organization), tragi<sub>A</sub>-comédi<sub>N</sub> (tragi-specifies the type of comedy), méga<sub>A</sub>lith<sub>N</sub> (mega- denotes the size of the architecture) ... The compound nouns of this type are neoclassical, in which N could be complex, but they select the structure of the compound nouns taken as inputs. They generally accept neoclassical attributive compound nouns NN(socio<sub>A</sub>[bionlogi<sub>N</sub>]<sub>CN</sub>) and AN (medico<sub>A</sub>-[socioalogi<sub>N</sub>]<sub>CN</sub>, poly<sub>A</sub> [glyco<sub>A</sub>lyse<sub>N</sub>]<sub>CN</sub>...).

#### 4.3.3. Subordinate Compound Nouns

Both native and neoclassical subordinate compound nouns of N1+N2 display the semantic relations for specifying the telic relation with respect to the head. For example, pause-café (café indicates the purpose of the break), wagon-fumeur (fumeur denotes the users of the wagon), bébé-éprouvette (éprouvette indicates the method of conceiving a baby), velodrome (vélo denotes that the path is constructed for bicycles), alcootest (alcool specifies test content of the analyzer), hydrothérapie (hydro- indicates the method of treatment), gynécologue (gynéco- denotes the field of the specialist) ... Like attributive compound nouns of NN, the productivity of this type of compound nouns is also very limited. N in the pattern N1+N2 is mainly simple and the compound noun NN is rarely used as an input for forming a new morphological compound.

The subordinate compound nouns of V+N are all native compounds. For neoclassical compound nouns, the order is reversed: N+V. This type of compound nouns replays the semantic relation of proto-patient, such as allume-cigar, appui-pouce, abat-jour... The compounds of V+N are very productive in native compounding, while most of the N+V compounds are adjectives (anglonphob<sub>V</sub>, mél<sub>N</sub>man<sub>V</sub>, noctinluque<sub>V</sub>...). The subordinate compound nouns of VN often serve as an input for forming a new subordinate native compound nouns of both NN and VN, for example, arbren [porte-galet]<sub>N</sub> and portev-[cure-dents]<sub>N</sub>. The combining ability of compound nouns VN is much less restricted in native compounding.

#### 4.3.4. The Other Types of Compound Nouns

The syntactical compound nouns can be divided into three types: N+Modifier (including N1+PREP+N2/V/VO, N1+N2, N+Pprés, N+Pp), V+Complement (V+PREP+N/A/Divers and VO) and V+et+V/A+et+A. In N+Modifier, the modifier specifies the function or property of the N, such as poche passepoilée (passepoilée denotes the appearance of the pocket), installateur en électroménager (en électro-ménager indicates the work field) ... For V+Complement, the complement is governed by verb and generally refer to the method (pince sans rire), result (porte à faux) or location (saute en bas) of the action or event. The complement in V+Complement can also be indirect object introduced by a preposition, for example, à tout in touche-à-tout. V+et+V and A+et+A are often composed of a pair of antonyms (aller et retour, lève et baisse, un noir et blanc, un chaud et froid...). The two adjectives describe a contrastive property or a state. Compounds of A+et+A are rarely taken as inputs for compounding new nouns. However, syntactical compounding is much more productive

than the morphological compounding. They can accept almost all types of morphological or syntactical compound nouns as inputs for forming new compound nouns. Even some non-productive patterns of compounds are accepted, for example, the non-productive compound adjective *nouveau-né* in the compound noun *enfant nouveau-né*.

## 5. Conclusion

This article firstly presented some problematic issues in definition of compound nouns, like separation of morphological from syntactical compounding, typology of compound nouns and delimitation of boundaries between derivation and compounding. Then, the discussion focused on constituents of French compound nouns: word, lexeme, stem, syntagm or modifier, and emphasized that the compound noun constituents could be complex units, in which there may be other complex units. For both morphological and syntactical compounding, it is not a simple word “lexeme” that can give a relevant definition for the notion “constituent”. Finally, it analyzed combining rules of French compound nouns, either morphological or syntactical, from the point of view of their phonological, categorical and semantic constraints. In this article, morphological and syntactical compounding are treated as two different formation processes of word, but both morphologically-formed and syntactically-formed nouns are considered as compound nouns. It is argued that morphological and syntactical constructions may be taken in each other despite the distinction between them. There is no distinct restriction of combination between them for forming a new compound noun, although syntactical compounds are more productive than native and neoclassical compounds. On the contrary, the exchanges between them are very frequent and productive in formation of compound nouns. The discussion about both constituents and combining constraints in this article make for better understanding formation process of French compound nouns, which contributes to both future linguistic research and studies in natural language processing.

## Acknowledgements

This research project is supported by Science Foundation of Beijing Language and Culture University (The Fundamental Research Funds for the Central Universities, item No.17YBB14).

## References

- [1]. Anderson, S. R. (1992). *A-morphous Morphology*. Cambridge: Cambridge University Press.
- [2]. Amiot, D., and Dal G. (2008). « Composition néoclassique en français et ordre des constituants », in D. Amiot éd. *La composition dans une perspective typologique*, 89-113. Arras, Artois Presses Université.
- [3]. Apothéloz D. (2002). *La construction du lexique français, Principes de morphologie dérivationnelle*. Paris: Ophrys.
- [4]. Arnaud, P. J. L. (2015). “Noun-noun compounds in French”, in: Müller P. O. et al. (eds), *Word-formation. An International Handbook of the Languages of Europe*, volume 1, 673-687.
- [5]. Benson M., Benson, E. and Ilson, R. (1986). *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Pays-Bas: John Benjamins.
- [6]. Booij, G. (2002). *The morphology of Dutch*. Oxford: Oxford University Press.
- [7]. Booij, G. (2005). « The grammar of words. An introduction to linguistic morphology ». Oxford : Oxford University Press.
- [8]. Corbin, D. (1992). « Hypothèses sur les frontières de la composition nominale ». *Cahiers de grammaire* 17, 25-55.

- [9]. Corbin, D. and Paul, J. (1999). « Aperçus sur la créativité morphologique dans la terminologie de la chimie ». *La banque des mots* 60, 51-68.
- [10]. Darmesteter, A. (1875). *Traité de la formation des mots composés dans la langue française comparée aux autres langues romanes et au latin*. France: A. Franck.
- [11]. Di Sciullo, A. M. (2005). “Decomposition compounds”. In *SKASE Journal of Theoretical linguistics* 2, 14- 33.
- [12]. Fradin, B. (2000). “Combining forms, blends and related phenomena”. In Ursula Dolsescha, Anna Maria Thrornton (eds), *Extragrammatical and marginal morphology*, 11-59. München: Lincom Europa.
- [13]. Fradin, B. (2003). *Nouvelles approches en morphologie*. Paris : Presses universitaires de France.
- [14]. Fradin, B. (2009). “Compounding in French”. In Lieber R. & P. tekauer(eds). *Oxford Handbook on Compounding*, 417-435. Oxford: Oxford University Press.
- [15]. Grevisse, M., 1986, *Le bon usage*, Belgique, Duculot.
- [16]. Gross, M. (1986). “Lexicon Grammar: The Representation of Compound Words”. In *Proceedings of Computational Linguistics (COLING’86)*, 1-6.
- [17]. Gross, M. (1996). *Grammaire transformationnelle du français : Syntaxe du nom*. France, Cantilène.
- [18]. Gross, G., Chaurand J., Mathieu-Colas M., Vives R. and Billy P. (1986). *Typologie des noms composés*. La Sorbonne Université Paris XIII.
- [19]. Guilbert, L. (1997). *La créativité lexicale*. Paris: Larousse.
- [20]. Haspelmath, M. (2002). *Understanding morphology*. London, Arnold.
- [21]. Huot, H. (2001). *Morphologie, forme et sens des mots du français*. France : A. Colin.
- [22]. Lesselingue, C. and Villoing F. (2002). *Aide-jardinier, aide-sage-femme, mots composés ou préfixés ?* In Peggy Afuta, Adil El Ghali & François Toussenet (eds.), *Actes de 7èmes rencontres de l’atelier des doctorants en linguistique (ADL’2002)*. Ecole doctorale de sciences du langage. La Sorbonne Université Paris VII.
- [23]. Lieber, R. (1992). “Deconstructing Morphology: Word Formation in Syntactic Theory”. University of Chicago Press.
- [24]. Mathieu-Colas M. (1996). « Essai de typologie des noms composés français ». *Cahiers de lexicologie*, 71-125. Didier Erudition.
- [25]. Matthes, Peter H. 1991. *Morphology. An introduction to the theory of word-structure*. 2nd edn. Cambridge: Cambridge University Press.
- [26]. Mithun, M. (2010). “Compounding and Incorporation”. In Scalise, S. & Vogel I. (eds.), *Handbook of Compounding*. John Benjamins.
- [27]. Montermini, F. (2010). “Units in compounding”. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, 77-92. Amsterdam & Philadelphia: John Benjamins.
- [28]. Norde, M. (2009). *Degrammaticalization*. Oxford: Oxford University Press.
- [29]. Riegel, M., Pellat, J. C. and Rioul, R. (1994). « *Grammaire méthodique du français* ». Coll. *Linguistique nouvelle*. Paris, Presses Universitaires de France.
- [30]. Saussure F. D. (1916). *Cours de linguistique générale*. France, Payot.

- [31]. Scalise, S. and Vogel, I. (2010). "Cross-Disciplinary Issues in Compounding". Amsterdam : John Benjamins.
- [32]. Van Goethem, K. (2012). « Le statut des séquences "N+N à N2 productif". Le cas de N-clé ». In: *Linguistica Investigations: revue internationale de linguistique française et de linguistique générale=international journal of French linguistics and general linguistics*, Vol. 35, no. 1, 76-93.
- [33]. Villoing, F. (2012). « French compounds ». *Probus, International Journal of Latin and Romance Linguistics*, 24(1), 29-60. Amsterdam: Mouton de Gruyter.