A paradigmatic account of lexical innovation: the role of repeated components in French N+N compounds

Jan Radimský
University of South Bohemia
radimsky@ff.jcu.cz

1. Introduction

The aim of this paper is to examine and refine, on both theoretical and empirical grounds, current hypotheses accounting for the emergence of the subordinate N+N compounding pattern in Romance that yields compounds such as the French *sauce-tomate* 'tomato sauce' and *exposition photos* 'photography exhibition'. More specifically, it will be argued, with reference to corpus data from French, that the emergence of new N+N compounds as well as the conventionalization of the N+N pattern is reinforced by the family-size effect of repeated nouns in either N1 or N2 position.

From the theoretical standpoint, the emergence of subordinate N+N compounds (henceforth SUB-NNs) in French will be accounted for in terms of Construction Grammar, as a consequence of two parallel processes: constructional change and constructionalization (Traugott & Trousdale 2013). From the former perspective, SUB-NNs can be viewed as non-conventional extensions of attributive N+N compounds (ATTR-NNs), such as *bourgeois gentilhomme* 'bourgeois gentleman', that have always been present in French (Radimský 2019a). The constructional change here consists in the fact that the semantic relationship between the components of left-headed NNs is no longer limited to attributive relations. From the latter perspective, it can be assumed that the increasing "critical mass" of single non-conventional SUB-NNs (including those that emerged as calques from Germanic languages, for example) led to a bottom-up constructionalization within the SUB-NN pattern.

The exact way the SUB-NN pattern has grown and the reasons that have encouraged its growth are, however, still far from being clear. Bearing in mind that almost any French SUB-NN may be expressed using the well-established N-PREP-N pattern, previous empirical research has focused primarily on the types of semantic relationship between the components of SUB-NNs, assuming that there should be some regularity and/or specificity which would allow to explain the growth of the SUB-NN pattern and, in general, its role among the processes of word-formation or naming. So far, this path has proved to be inconclusive: as pointed out by Arnaud (2003: 64), the French SUB-NN pattern does not seem to have any restrictions nor general preferences as for the semantic relationship between the components, and the lack of any semantic regularity is "frustrating". However, empirical data from the Frantext corpus show that the SUB-NN pattern as such began to grow exponentially, in terms of both type and token frequency, during the 2nd half of the 20th century (Radimský 2019a), which implies that there must be some regularity behind this process. The hypothesis examined in this paper states that the regularity that underpins the exponential growth of new SUB-NNs during the 20th century is based on the progressive constructionalization of semi-schematic daughter constructions within the SUB-NN pattern. In other words, it will be argued that the spread and

MMM12 Online proceedings 2020
J. Audring, N. Koutsoukos & C. Manouilidou (Eds)

conventionalization of the N+N pattern from the 1960s onwards is reinforced by a family-size effect of repeated components, be it in the left-hand or the right-hand position.

The paper is organized as follows: Section 2 presents an outline of initial assumptions. In Section 3 the data from Frantext and FrWac corpora will be discussed from a quantitative point of view, introducing component-based regularities. Some nascent general sense-based regularities of the SUB-NN pattern that emerge from the data will be sketched in Section 4.

2. A CM model of emergence of SUB-NNs

Romance SUB-NNs appeared progressively during the 19th-20th centuries as instances of a new pattern that enters competition with older and extremely profitable patterns of phrasal lexemes (N+PREP+N and N+A). From a theoretical standpoint, there is still no agreement as to the causes that led to the emergence and spread of this pattern. In the literature on the topic, the following four main hypotheses are discussed (see Arnaud 2003): (a) an Old French N+NOBL structure; (b) the instability of prepositions in N-PREP-N structures; (c) the influence of Germanic languages; and (d) analogy to attributive N+N structures. The first hypothesis (a) has already been refuted (Rainer & Buridant 2015), and it has also been shown that the factors (b) and (c) had a relatively marginal effect (Picone 1996; Arnaud 2003). Indeed, while the influence of English (c) has probably contributed to the spread of some SUB-NNs in the 2nd half of the 20th century, it cannot explain the emergence of older SUB-NNs (Arnaud 2003: 138) and, thus, the emergence of the SUB-NN pattern as such. Conversely, hypothesis (d) that conceives Romance SUB-NNs as a direct continuation of ATTR-NNs deserves more attention. The idea has been put forward en passant in a footnote by Darmesteter (1874: 138-139) who noticed that in his times the ATTR-NN pattern was becoming a kind of "mould" which, by analogy, also embraced N+N compounds with a subordinate relationship between the constituents. Darmesteter's intuition has recently been reformulated in terms of Construction Morphology (Radimský 2019a), with reference to the notion of a hierarchical organization of schemas (Booij 2016), as will be briefly outlined in the following paragraphs.[1]

In terms of the compound typology put forward by Scalise and Bisetto (Bisetto & Scalise 2005, and Scalise & Bisetto 2009), contemporary French has three major types of N+N compounds, i.e. Attributive, Coordinate and Subordinate compounds, exemplified in (1a), (2) and (3), respectively (see also Villoing 2012). Type (1b) with a metaphoric interpretation of the non-head noun has a debatable status: it has been classified either as attributive (Bisetto & Scalise 2005), subordinate (Arnaud 2003), or as a specific "appositive" subtype of a larger attributive-appositive group (Scalise & Bisetto 2009; Radimský 2015), which will also be the approach adopted here.

(1) a. Attributive N+N compound (*bourgeois gentilhomme* 'bourgeois gentleman')
$[N_iN_j]_{Nk} \leftrightarrow [SEM_i \text{ is a } SEM_j]_k$

b. Appositive N+N compound (*oiseau-mouche* 'hummingbird')
$[N_iN_j]_{Nk} \leftrightarrow [SEM_i \text{ is like a } SEM_j]_k$

(2) Coordinate N+N compound (*libraire-éditeur* 'bookseller-publisher')
$[N_iN_j]_{Nk} \leftrightarrow [SEM_i \text{ is a } SEM_j]_k \leftrightarrow [SEM_j \text{ is a } SEM_i]_k$

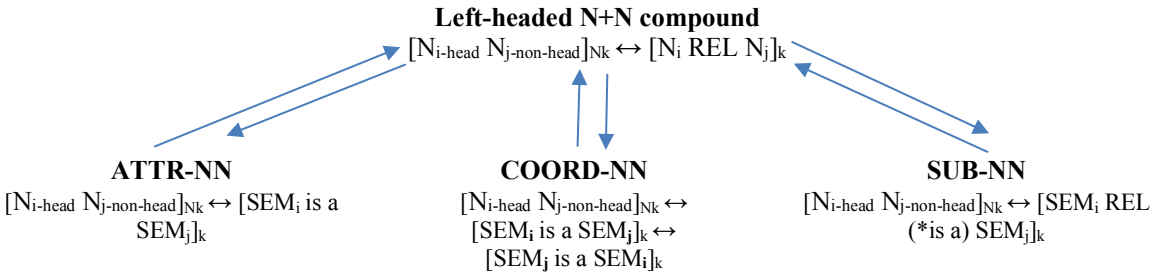
(3) Subordinate N+N compound (*timbre-poste* 'stamp')
$[N_iN_j]_{Nk} \leftrightarrow [SEM_i \text{ REL(*is a) } SEM_j]_k$

[1] A different formalization of this hypothesis put forward by Hatcher (1946: 216-217), which consists in the idea of a progressive logical extension of the different semantic relationships between the N+N components, proved not to be compatible with the empirical data.

Within the Construction morphology approach (CM, Booij 2010, 2016), each of these types may be conceived as a construction formalizable by a schema, as indicated above. The abbreviation "REL(*is a)" in the schema that captures SUB-NNs (3) stands for any type of semantic relationship except the attributive one. Notice that COORD-NNs (2) are mere "symmetric" variants of ATTR-NNs (1a),[2] which, by the way, supports the hypothesis put forward by Rainer and Buridant (2015: 1978) that, diachronically, COORD-NNs derive from ATTR-NNs.

An important aspect of the CM approach is that constructions are organized hierarchically (Booij 2016: 430-433), so that lower-level constructions that share some common properties may be subsumed under a more abstract higher-level construction and, in the opposite perspective, lower-level constructions share or inherit the properties of higher-level constructions. Thus, constructions capturing the different subtypes of French NNs (1-3) are instances of a more general left-headed N+N construction, as indicated in Fig. 1.[3]

Figure 1: Hierarchy of French N+N constructions



This model allows us to hypothesize that not only coordinate (Rainer & Buridant 2015: 1978) but also subordinate and appositive NNs represent "non-conventional extensions"[4] of the original attributive pattern in the sense that the attributive "is a" relation is extended to any semantic relation ("REL"). To put it differently, COORD-NNs and SUB-NNs would result from a "constructional change" of the ATTR-NN pattern in the sense of Traugott and Trousdale (2013). Empirical support to this hypothesis is given in Radimský (2019a), where type and token frequency curves of ATTR-, COORD-, and SUB-NNs in the Frantext corpus are compared for the period of 1830-1999. This data shows that the diachronic rise of the frequency curves of ATTR and COORD-NNs matches that of the curves of SUB-NNs: they are almost stable for more than a 100 years, but an exponential increase has been observed since the 1960s.

The hypothesis outlined above still leaves some important questions open. There is no doubt that the COORD-NN pattern represents a construction in the sense of CM, because when instantiated by any pair of nominal components, it yields a well-acceptable compound in contemporary French.[5] However, the same does not hold for the SUB-NN pattern, as described

---

[2] Indeed, the constituents of COORD-NNs may be inverted, which potentially entails a slight semantic shift (for a detailed discussion, see Radimský 2015: 102-112);

[3] The appositive subtype (1b) is excluded from Fig. 1 for reasons of space.

[4] "... [the] tolerance for nonconventionality is of great importance in change: partially sanctioned extensions of an existing conventionalized construction may over time become fully sanctioned instances of a more general, schematic construction, which has changed as a result of the speaker/hearer's experience with language" (Traugott & Trousdale 2013: 16)

[5] Notice, however, that components of COORD-NNs must be semantically related and situated on the same level of a semantic hierarchy, such as, for instance, two nouns denoting professions. Otherwise the compound will necessarily have an asymmetrical, attributive interpretation (for a detailed discussion see Radimský 2015: 102-112).

in Fig. 1. On the one hand, the schema characterizing SUB-NNs in Fig. 1 captures several hundreds or thousands contemporary French NN compounds, so it allows speakers to correctly interpret novel NN types they encounter. But on the other hand, the same schema is not instantiated freely on an open-choice basis: in many cases, the resulting compound would sound inacceptable or odd. This situation raises two interesting questions: what is the status of the construction capturing French SUB-NNs and how can we explain the exponential growth of type and token frequency of French SUB-NNs since the 1960s?

A plausible explanation must take into account the fact that the constructional change, initiated more than two centuries ago, is still unterway. Until the 1960s, the schema of SUB-NNs in Fig. 1 was probably nothing more than a non-conventional extension of the attributive pattern that helped the linguistic community to form and accept sparse new instances of SUB-NNs, as noticed by Darmesteter. So, the vocabulary of SUB-NNs kept growing very slowly for over a 100 years, until a critical mass of instances had formed by the middle of the 20th Century. At this point, subsets of SUB-NNs with similar properties progressively began to encourage further bottom-up constructionalization that lead to the creation of new specific "daughter" SUB-NN constructions. Therefore, the exponential growth of the SUB-NN pattern, observed since 1960s, may be due to the rise of these new lower-level daughter constructions. If this explanation is correct, we should be able to identify these daughter constructions of the SUB-NN pattern in the data.

It might seem reasonable to assume that lower-level SUB-NN constructions would be based on the same semantic relationship between the compound components, i.e. on a specified "REL". However, previous research has shown that, surprisingly, this does not seem to be the case. Arnaud (2003: 64) has identified no less than 54 different types of semantic relations ("RELs") in a sample of nearly 1000 French NNs, and he concludes that the lack of semantic regularity is frustrating. On the one hand, many different semantic relations – including many complex ones – may be found within French SUB-NNs but, on the other hand, there is probably not a single one REL that would yield new fully acceptable SUB-NN compounds on an open-choice basis, i.e. using any semantically appropriate pair of nominal components. Even the so-called "purpose" REL, which is reported to be the one with the highest type frequency within Arnaud's sample (247 types!),[6] does not yield perfectly acceptable NNs with just any component, as indicated in (4a-b).

(4) $[N_iN_j]_{Nk} \leftrightarrow [SEM_i \text{ is for } SEM_i]_k$

a. *bâtiment voyageurs* ('passenger building') (Arnaud 2003: 98)
b ?*hôpital enfant(s)* ('children's hospital')[7]

Therefore, postulating daughter constructions with a specified REL, such as the one in (4a), does not really help to explain the massive creation of novel SUB-NN compounds since the 1960s.

An alternative hypothesis consists in arguing that the increasing productivity of the SUB-NNs during the 2nd half of the 20th century is underpinned by a component-based regularity, i.e. by the family-size effect of repeated components in either the left- or the right-hand position.

---

[6] In Arnaud (2003: 65), the "purpose" REL is described as follows: "N2 est la destination, l'objet de N1 (général)". It is, however, still a macro-class that encompasses different types of semantic relations under a more fine-grained view.

[7] The sequence *hôpital enfant(s)* has only 7 hits in the FrWac corpus, all related to one hospital (Timone in Marseille).

So far, the role of repeated components in Romance N+N compounds has only been studied with reference to ATTR-NNs (Baroni, Guevara & Pirrelli 2009; Grandi 2009; Radimský 2015, amongst others) because N2s with a specific meaning (those that express "quality") trigger the attributive interpretation of the whole N+N compound. In such cases, component-based regularity is directly linked with sense-based regularity, i.e. with the specified attributive "REL". However, the analysis by Baroni, Guevara, and Pirrelli (2009) has also revealed that, surprisingly, even in Italian SUB-NNs there are some typical N2s with a high type frequency, such as *sicurezza* 'security'*, stampa* 'press', and *lavoro* 'work', though there is no plausible semantic rationale lying behind such regularities. Such nouns occur frequently in the right-hand position of Italian SUB-NNs, but in a wide variety of semantic relations to their left-hand members, and sometimes even with different meanings if they are polysemous.[8] Similarly, Koga (2018) has recently identified some N1s and N2s with a significant type frequency in French SUB-NNs. In CM terms, this can be explained by the fact that the set of previously attested NNs with the same item on a given position yields a semi-schematic construction, such as (5a), which in turn allows users to form new NNs (types) based on the same form, irrespective of the semantic relationship (REL) between the components.

(5) a. $[N_i\ stampa_j]_{Nk} \leftrightarrow [SEM_i\ REL\ press_j]_k$ *conferenza stampa* ('press conference')
b. $[N_i\ stampa_j]_{Nk} \leftrightarrow [SEM_i\ REL\ SEM[stampa]_j]_k$ *centro stampa* ('printing center')

Notice that in (5a), the construction reflects only one meaning of the polysemous noun *stampa*, which happens to be the most frequent one in SUB-NNs; indeed, it is even mentioned by the Zingarelli (2011) dictionary as the specific bound meaning of *stampa* in NNs. Nevertheless, it is not impossible to find examples of N2s that reflect the original polysemy of the component, such as in (5b).

This hypothesis predicts that within a set of French NNs it should be possible to identify some patterns (i.e. subsets of types) based on repeated N1s or N2s, irrespective of the question whether the semantic REL between the two nouns is identical or not. In the next section, it will be shown that this is the case.

3. Data analysis

The hypothesis outlined above was tested on data samples from two corpora, Frantext and FrWac, respectively, which also allowed us to make a basic diachronic comparison between the two datasets. The next two paragraphs will briefly introduce the data-gathering process, then a detailed analysis of frequency spectra will follow.

Frantext is a large corpus (127M tokens) containing mainly French fiction, and its lemmatized section covers texts from between 1830 and 1999. From this corpus, complete concordances of binominals in both hyphenated and non-hyphenated form were extracted. Subsequent lemmatization made it possible to identify 33,800 binominals (lemmatized types). A manual filtering lead to the identification of a sample of 1,631 N+N compounds comprised of 299 SUB-NNs and 1274 ATTR/COORD-NNs. This data is presented in Radimský (2019a), where the extraction procedure is described more thoroughly. In the present paper, we focus

[8] Frequent SUB-NNs with *stampa* refering to 'press, media' in the ItWac corpus are, for example: *addetto stampa* 'press agent', *conferenza stampa* 'press conference', *rassegna stampa* 'press review' or *silenzio stampa* 'media blackout'; the respective paraphrases of the REL would necessarily differ. Another frequent compound, *centro stampa*, may mean 'press center', but also 'printing center', where the component *stampa* refers to the 'action of printing'.

primarily on the repeated forms in SUB-NNs. This is why the initial set of 299 SUB-NNs has been rechecked, and new SUB-NNs based on repeated forms were retrieved from the database, so that the number of SUB-NNs in the sample increased to 353 types. Since the Frantext data was intended to show the role of repeated forms before the 1960s, the sample of SUB-NNs was filtered again in order to include only types for which the average year of appearance in Frantext does not exceed the year 1959. Thus, the final dataset of "old" SUB-NNs from Frantext is comprised of 152 types.

FrWac is a very large web corpus (1600M tokens) developed between 2005 and 2007. It contains mainly texts written around the year 2000 and pertaining to various genres. As in the case of Frantext, it served as a basis for the extraction of frequency lists of binominals (430,000 types with a minimum token frequency set to 3), out of which a sample of 3,350 N+N compounds was identified manually. Again, manual sampling from previous research was used (Radimský 2018a, 2018b), but the data was rechecked, and new types based on repeated components were added. The final sample of "modern" NNs comprises 2,112 SUB-NNs and 1,238 ATTR-NNs, respectively.

To sum up, the analysis will focus on three different datasets: "old" SUB-NNs from Frantext (152 types), "new" ATTR-NNs from FrWac (1238 types), and "new" SUB-NNs from FrWac (2112 types). For each dataset, two frequency spectra were calculated in order to show the profile of repeated lemmas in the left-hand (N1) and right-hand (N2) position, respectively (see Tables 1-6; Tables 3-6 can be found in the Appendix). In these datasets it will be particularly interesting to compare the rate of components with type frequency f=1 (see the first column of Table 7 for an overview): higher rates indicate a smaller effect of repeated components.

Table 1: Frequency spectrum of N1s in "old" SUB-NNs (Frantext)[9]

f	V(f)	V(f) rel.	Examples of N1s
1	99	84.6%	
2	8	6.8%	
3	7	6.0%	service, wagon, assurance 'insurance', café 'coffee'
4	1	0.9%	homme 'man'
6	2	1.7%	question, papier 'paper'

Table 2: Frequency spectrum of N2s in "old" SUB-NNs (Frantext)

f	V(f)	V(f) rel.	Examples of N2s
1	83	83.8%	
2	8	8.1%	
3	1	1.0%	
4	3	3.0%	
5	1	1.0%	santé 'health'
6	1	1.0%	poste 'post office'
9	1	1.0%	frontière 'border'
18	1	1.0%	maison 'home-made', lit. 'house'

Frequency spectra for "old" SUB-NNs from Frantext, presented in Tables 1 and 2, show that the effect of repeated nouns is very weak. In both positions (N1 and N2), the rate of nouns with

[9] f = type frequency (number of NNs with the same N1 in the sample)
V(f) = number of nouns (N1s) in the sample with type frequency f
V(f) rel. = proportion of nouns (N1s) in the sample with type frequency f

the lowest type frequencies (f=1) is very high –around 84%– and there are also very few nouns with a higher type frequency. Thus, in CM terms, most of these NNs are created as individual non-conventional extensions of attributive NNs that fit the general left-headed N+N pattern schematized as $[N_{i\text{-head}}N_{j\text{-non-head}}]_{Nk} \leftrightarrow [\text{SEM } N_i \text{ REL SEM } N_j]_k$, while only a few of these NNs already yield a more specific semi-schematic construction with a lexically filled N1 or N2 and, in some cases, even with a specified semantic relationship. The semantic regularity holds true, for instance, for the noun *maison* ('house') that in the N2 position systematically appears with the bound meaning 'home-made' which, in turn, happens to be so specific that it systematically correlates with the same semantic REL to the left-hand member of the compound.

Frequency spectra for "new" ATTR-NNs from FrWac are presented in Tables 3 and 4 (see Appendix). As expected, the frequency effect of repeated nouns is very strong at the N2 position (Table 4), where only 30% of N2s has the type frequency f=1, and many N2s have higher type frequencies.[10] The opposite can be seen in the case of N1s (Table 3). Although the rate of nouns with the lowest type frequency (f=1) is much lower than for "old" data from Frantext (67.1% vs. 84%), the list of nouns with higher type frequencies shows that the effect of repeated nouns is very weak. Indeed, in 85.1% of types the same N1 appears twice at most,[11] and the N1 with the highest type frequency *ville* 'town' yields only 15 ATTR-NN types.

The empirical support for the hypothesis discussed in Section 2 is visible on frequency spectra of "new" SUB-NNs from FrWac, presented in Tables 5 and 6, respectively (see Appendix). This data shows that there are "typical" N1s as well as "typical" N2s with a high type frequency: 32 N1s and 39 N2s appear in more than 9 SUB-NN compounds (types). The list of 15 nouns with the highest type frequency for the N1 and N2 position is given in (6) and (7), respectively.

(6) **N1s with the highest type frequency in new SUB-NNs**

pôle 'pole', rayon 'department', secteur 'sector', atelier 'workshop', service 'service' , papier 'paper', soirée 'evening', rubrique 'rubric', accès 'access', coin 'corner', assurance 'insurance', filière 'branch', sauce 'sauce', association 'association', guide 'guide', espace 'space'

(7) **N2s with the highest type frequency in new SUB-NNs[12]**

radio 'radio', client 'customer', jeunesse 'youth', papier 'paper', achat 'purchase', auto 'car', maison 'home-made' (lit. 'house, home'), bébé 'baby', santé 'health', zen 'calm' (lit. 'zen'), bidon 'fake' (lit. 'can'), beauté 'beauty', photo 'photo', radar 'radar', aluminium 'aluminum'

The comparison of the rate of components with type frequency f=1 for all the different datasets is given in Table 7. Provided that a lower rate of components with type frequency f=1 indicates a higher number of larger families with the same constituent, these figures show that from the diachronic point of view, the effect of component repetition in SUB-NNs increased. Notice also

[10] Besides, notice that the N2 with the highest type frequency *clé* 'key' yields 220 NN types, which represents almost 18% of all ATTR NNs (types) in the sample (the dataset of new ATTR-NNs contains 1,238 types, so 220/1,238=17.78%).

[11] These 85.1% represent the sum of 67.1% of N1s with f=1 and 18% N1s with f=2 (Tab. 3).

[12] The nouns *maison, zen* and *bidon* appear with a specific bound meaning which differs from their free meaning.

that in the case of N1s, the family effect in new SUB-NNs is even higher compared to new ATTR-NNs.

Table 7: Comparison of different frequency spectra

	Rate of Ns with type fq.=1	
	N1	**N2**
Old SUB-NNs	84.6%	83.8%
New SUB-NNs	58.3%	61.7%
New ATTR-NNs	67.1%	30.4%

It terms of CM, it can thus be argued that the sample of new SUB-NNs comprises a set of new semi-schematic constructions, based on either a specified N1 or a specified N2 (see examples of the respective components in 6-7 above), which result from the process of bottom-up constructionalization. These "daughter" constructions of the general SUB-NN construction are very probably responsible for the exponential growth of the SUB-NN pattern, observed since the 1960s.

It might seem surprising that in the sample of new SUB-NNs, the effect of repeated nouns is visible and almost balanced in both positions. Indeed, one may assume that the head could be of more importance for the semantic relation between the compound members than the non-head, so that the family effect of repeated items should be higher for N1s.[13] However, this is not the case for two reasons. On the one hand, the different head nouns (N1s) are not systematically linked to the same REL (see examples 10a-d in the next section) and sometimes they are even polysemous: for instance, the noun *rayon* appears as head noun in SUB-NNs not only with the meaning 'department, section of a department store', but also with the meaning 'ray, beam'. On the other hand, some N2s such as *maison, zen* or *bidon* (cf. 7) tend to appear with a specific bound meaning which is linked to the same REL. In other words, the semantic cue in SUB-NNs may be related either to the head noun or the non-head noun.

4. General semantic regularities within the SUB-NN pattern

The data analyzed in Section 3 makes it possible to claim that as the vocabulary of SUB-NNs progressively grows, there is more regularity based on repeated N1s and N2s. Such regularity is component-based, which means that in CM terms, a bottom-up constructionalization yields semi-schematic constructions with a lexically filled N1 or N2, such as (8) and (9).

(8) $[papier_{i}N_{j}]_{Nk} \leftrightarrow$ ['paper' REL $SEM_{j}]_{k}$
(9) $[N_{i}client_{j}]_{Nk} \leftrightarrow [SEM_{i}$ REL 'client'$]_{k}$

It is interesting to notice that the semantic relationship REL in constructions like (8-9) often remain underspecified, unless the fixed component has a specific bound meaning that triggers a specific REL (see examples in 7). In other words, the same N1 or N2 is not necessarily associated with the same REL. Examples of different semantic RELs for the semi-schematic construction (8) are given in (10a-d).

(10) a. *papier journal* ('newsprint' = low-cost paper commonly used to print newspapers, lit. 'newspaper paper'): N2 is an object typically made of the kind of N1

[13] Notice that this principle holds for ATTR-NNs, where the figures are not balanced, because the semantics of non-head items is usually correlated with the attributive REL.

b. *papier aluminium* (‘aluminium foil’, lit. ‘aluminium paper”): N2 specifies the material of N1
c. *papier cuisson* (‘baking paper’): N2 is an event that N1 allows to perform
d. *papier cadeau* (‘wrapping paper’, lit. ‘gift paper”): N2 is an object affected by the event that N1 allows to perform

Nevertheless, this does not mean that further bottom-up processes of constructionalization based on purely semantic generalizations would be impossible. Indeed, the data suggests that it is beginning to take place on two different levels, either as a semantic generalization over *one* semi-schematic construction, or as a generalization over *a set* of different semi-schematic constructions. Both processes will be briefly illustrated in the following paragraphs.

Let us first focus on the process of semantic generalization over one semi-schematic construction. Although we may expect it to consist in a simple specification of the semantic REL between the components, data from our sample of SUB-NNs shows that it is not always so straightforward. Examples (11a-e) show five different SUB-NNs that instantiate the same semi-schematic construction (11). Compared to (8), the construction (11) is not only component-based, since the compounds in (11a-e) have some common semantic “denominator” – they all denote different sections of a department store (see also Koga 2018).

(11) [*rayon* $N_j]_{Nk}$ ↔ [‘department’ REL $SEM_j]_k$
a. *rayon boucherie* (‘meat counter’, lit. ‘butchery department’)
b. *rayon enfant* (‘children’s section’)
c. *rayon jouets* (‘toys department’)
d. *rayon surgelés* (‘[frozen food] section‘)
e. *rayon* [*fruits et légumes*] (‘[fruits and vegetables] department’)

Notice, however, that these “sections” are referred to from three different perspectives, so that the N2 may denote the names of the respective stores (11a), the target customers (11b), or the goods sold at the respective departments (11c-e). To put it differently, the identity of the semantic REL is intuitively straightforward for compounds (11c-e), but there is clearly also a more abstract –and perhaps more important– semantic REL underlying the whole set (11a-e) that may be schematized as (12).

(12) [*rayon*$_i$ $N_j]_{Nk}$ ↔ [section of a department store having SEM_j as salient feature]$_k$

Such more abstract REL does not imply that the semantic relationship between the compound’s components (N1 and N2) be the same. Its pertinence, however, is determined by the fact that the rightmost noun represents an open-choice slot, and the resulting compounds may be perceived as a neutral and conventional way to refer to the internal organization of a department store, irrespective of the way in which the REL between the components could be analysed on a more fine-grained level.

Concerning the second level of generalization, i.e. that over a set of semi-schematic constructions, it seems that the “critical mass” of the different semi-schematic constructions is still too small to allow systematic constructionalization. The best example are probably semi-schematic constructions similar to (12), based on different N1s as exemplified in (13a-g).

(13) a. *pôle+N2* (‘center’) – organization specialized in one activity (type fq. = 81)
b. *rayon+N2* (‘department’) – section of a department store (type fq. = 63)
c. *atelier+N2* (‘workshop’) – workshop specialized in one activity (type fq. = 42)
d. *service+N2* – service specialized in one activity (type fq. = 40)

e. *rubrique+N2* – (‘column’) newspaper section (type fq. = 35)
f. *coin+N2* – (‘corner, area’) area of a house or store dedicated to an activity (type fq. = 25)
g. *filière+N2* – (‘industry’) industry specialized in one activity (type fq. = 22)

All these N1s have a rather high type frequency, and the logic lying behind the common semantic REL consists in the fact that the respective compounds denote a conceptual subtype or subclass of the concept denoted by N1.[14] The set of semi-schematic constructions (13a-g) makes it possible to posit a more abstract schematic construction (14), which allows the N1 “slot” to be filled by any semantically acceptable noun on an open-choice basis. Indeed, a look in the FrWac corpus confirms that there are other possible N1s that fit the construction (14), such as *branche* ‘branch, field’, *espace* ‘space, area’, and *point* ‘point’.

(14) $[N_iN_j]_{Nk} \leftrightarrow$ [specific subdomain of SEM_i related to $SEM_j]_k$

Another example of the generalization over a set of semi-schematic constructions is given in (15). It is worth mentioning especially because the abstraction is based not only on semi-schematic constructions with a repeated N1, such as *sauce* (as in *sauce tomate* ‘tomato sauce’) and *glace* (as in *glace vanille* ‘vanilla ice-cream’), but at the same time also on semi-schematic constructions with a repeated N2, especially in the case of sweet foods. Thus, the N2 *vanille* (‘vanilla’) yields different NNs based on the same semantic pattern, such as *sucre vanille* (‘vanilla sugar’), *rhum vanille* (‘vanilla rum’), *crème vanille* (‘vanilla cream’), and *yaourt vanille* (‘vanilla yogurt’), which helps to fix the construction (15).

(15) $[N_iN_j]_{Nk} \leftrightarrow$ [food SEM_i with flavour or ingredient specified by $SEM_j]_k$

Even in this case, the schematic construction (15) allows for different N1s and N2s on a more or less open-choice basis, at least for desserts (see N1s, such as *tarte* ‘pie’*, flan* ‘custard’, and *fondant* ‘fondant’). In this case, however, the productivity is limited by the fact that the construction in (15) is in direct competition with the conventionalized NPN construction (16), as illustrated in examples (17a-b).

(16) $[N_i$ *à* DET $N_j]_{Nk} \leftrightarrow$ [food SEM_i with flavour or ingredient specified by $SEM_j]_k$

(17) a. *glace vanille* (‘vanilla ice-cream”)
b. *glace à la vanille* (‘vanilla ice-cream”)

This brings us again to the more general question of the competition between the NN and the NPN pattern in French, which would certainly deserve deeper investigation.

5. Conclusion

Following Traugott and Trousdale (2013: 1), a constructional model of language change involves two types of phenomena: changes that affect existing constructions (called “constructional changes”), on the one hand, and creation of new form-meaning pairings (refered to as “constructionalization”), on the other. According to a different perspective (Traugott &

[14] See also Fradin (2003: 203) for this function of NN compounds.

Trousdale 2013: 21), a constructionalist model must account for both "innovations that apply to particular internal dimensions of a construction" and the "conventionalization of those innovations among a group of speakers". In this paper it has been shown how such a CM model of language change accounts for the emergence and the conventionalization of French subordinate NNs, focusingon the specific role of the family-size effect of repeated components in either N1 or N2 position in the process of constructionalization.

We started out from the assumption that the initial creation of French coordinate and subordinate NNs during the 19th and the first half of the 20th centuries was primarily a result of *constructional change* based on individual non-conventional innovations of the left-headed attributive NN pattern. Such "tolerance for nonconventionality" (Traugott & Trousdale 2013: 16) entails that in this period, it was possible to form any left-headed NN, but most of subordinate NNs were perceived as non-conventional. Indeed, conventionalization concerns individual micro-constructions only, such as *appareil photo* 'camera', *papier carbone* 'carbon paper', and *timbre-poste* 'stamp'.

The second stage of the process, which took place especially from the 1960s onwards when an exponential increase in productivity of French SUB-NNs is observed, consisted in a progressive bottom-up *constructionalization*, which was primarily component-based. That is, repeated nouns in either N1 or N2 position in the existing micro-constructions lead to the creation of semi-schematic constructions based on a lexically specified N1 or N2. From the empirical point of view, type frequency spectra changed (i.e. the rate of N1s and N2s with the lowest type fq. decreased), but semantic regularities within the global SUB-NN pattern were still extremely weak (Arnaud 2003), because a consistent abstraction of semantic patterns was linked at most to some fixed components of the semi-schematic constructions.

The examples analysed in Section 4 allow us to argue that more abstract and sophisticated sense-based constructionalization is also beginning to take place, even over sets of semi-schematic constructions. For the time being, French data from the FrWac corpus provides only a few examples, but Italian verbal-nexus NNs (see Radimský 2015, 2018b) may represent a good example of a new Romance NN pattern that is not purely component-based.

Analyses of new sense-based schematic SUB-NN constructions also clearly show that any future research in the domain will have to tackle seriously the issue of competition between subordinate NNs on the one hand and N-PREP-N or N-A structures on the other, since the latter represent firmly established patterns in French (cf. Radimský 2019b).

Acknowledgements

This research has been supported by the research grant GAČR (Grant Agency of the Czech Republic) No. GA17-17253S – "N-N Compounding in Contemporary French."

References

Arnaud, P. J. L. 2003. *Les composés timbre-poste*. Lyon: Presses universitaires de Lyon.

Baroni, M., E. Guevara & V. Pirrelli. 2009. Sulla tipologia dei composti N+N in italiano: princìpi categoriali ed evidenza distribuzionale a confronto. In R. Benatti, G. Ferrari & M. Mosca (Eds.), *Linguistica e modelli tecnologici di ricerca* (Atti del 40esimo Congresso della Società di Linguistica Italiana). Roma: Bulzoni, 73-95.

Bisetto, A. & S. Scalise. 2005. The classification of compounds. *Lingue e Linguaggio* 4(2): 319-332.

Booij, G. 2010. *Construction Morphology*. Oxford: Oxford University Press Booij, G. 2016. Construction Morphology. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology*. Cambridge: Cambridge University Press, 424-448.

.

Darmesteter, A. 1874. *Traité de la formation des mots composés dans la langue française comparée aux autres langues romanes et au latin*. Paris: A. Franck.

Fradin, B. 2003. *Nouvelles approches en morphologie*. Paris: PUF.

Frantext, Corpus de langue française. < https://www.frantext.fr/> (consulted March-April 2018)
FrWac - Baroni, M. et al. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3): 209-226.
Grandi, N. 2009. When Morphology 'Feeds' Syntax: Remarks on Noun > Adjective conversion in Italian Appositive Compounds. In F. Montermini, G. Boyé, J. Tseng (Eds.), *Selected Proceedings of the 6th Décembrettes*. Somerville, MA: Cascadilla Proceedings Project: 111-124.
Hatcher, A. G. 1946. Le Type Timbre-Poste. *Word* 2(3): 216-228.
Koga K. 2018. Un espace papeterie n'est-il pas une (simple) papeterie? Composition binominale souscatégorisante. *SHS Web Conf.*, Vol. 46, 08004, Congrès Mondial de Linguistique Française - CMLF 2018.
Picone M. D. 1996. *Anglicisms, Neologisms and Dynamic French*. Amsterdam: John Benjamins.
Radimský, J. 2015. *Noun+Noun Compounds in Italian. A corpus-based study*. České Budějovice: University of South Bohemia.
Radimský, J. 2018a. Inflection of binominal ATAP compounds in French and Italian: a paradigmatic account. *Lingue e linguaggio* XVII(2): 261-272.
Radimský, J. 2018b. Does French have verbal-nexus Noun+Noun compounds? *Linguisticae Investigationes* 41(2): 214-224.
Radimský, J. 2019a. Les composés N-N de subordination: un paradigme émergent. *Studia Romanica Posnaniensia* 46(1): 167-180.
Radimský, J. 2019b. Are French NNs variants of N-PREP-N constructions? Paper presented at *The second International Symposium of Morphology* (ISMo 2019), University Paris-Diderot, Paris, September 25-27, 2019, < http://drehu.linguist.univ-paris-diderot.fr/ismo-2019/>.
Rainer, F. & C. Buridant 2015. From Old French to Modern French. In: P. O. Müller, I. Ohnheiser, S. Olson, F. Rainer (Eds.), *Word-formation: an international handbook of the languages of Europe*. Vol. 3. Berlin/Boston: Walter de Gruyter, 1975-2000.
Scalise, S. & A. Bisetto. 2009. The classification of compounds. R. Lieber & P. Štekauer (Eds.), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 34-53.
Traugott, E. C. & G. Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
Villoing F. 2012. French compounds. *Probus* 24(1): 29-60.

Appendix

Table 3: Frequency spectrum of N1s in "new" ATTR-NNs (FrWac)

f	V(f)	V(f) rel.	Examples of N1s
1	496	67,1%	
2	133	18,0%	
3	46	6,2%	
4	31	4,2%	
5	14	1,9%	
6	10	1,4%	
7	3	0,4%	
8	1	0,1%	region 'region'
9	1	0,1%	site 'site'
10	2	0,3%	image 'image', zone 'area'
11	1	0,1%	pays 'country'
15	1	0,1%	ville 'town'

Table 4: Frequency spectrum of N2s in "new" ATTR-NNs (FrWac)

f	V(f)	V(f) rel.	Examples of N2s
1	42	30,4%	
2	23	16,7%	
3	8	5,8%	
4	6	4,3%	
5	9	6,5%	
6	3	2,2%	
7	1	0,7%	
8	6	4,3%	
9	4	2,9%	
10	2	1,4%	
11	3	2,2%	
12	2	1,4%	
13	3	2,2%	
14	2	1,4%	
15	3	2,2%	
16	2	1,4%	
17	1	0,7%	
18	4	2,9%	
20	1	0,7%	
21	1	0,7%	
22	2	1,4%	
24	1	0,7%	
26	2	1,4%	
27	1	0,7%	expert 'expert'
34	1	0,7%	surprise 'surprise'
36	1	0,7%	fantôme 'ghost'
49	1	0,7%	limite 'borderline'
53	1	0,7%	victime 'victim'
58	1	0,7%	membre 'member'
220	1	0,7%	clé 'key'

Table 5: Frequency spectrum of N1s in "new" SUB-NNs (FrWac)

f	V(f)	V(f) rel.	Examples of N1s
1	444	58,3%	
2	142	18,6%	
3	58	7,6%	
4	37	4,9%	
5	18	2,4%	
6	13	1,7%	
7	10	1,3%	
8	4	0,5%	
9	4	0,5%	
10	2	0,3%	
11	4	0,5%	
13	5	0,7%	
14	5	0,7%	
16	1	0,1%	espace 'space'
19	1	0,1%	guide 'guide'
21	1	0,1%	association 'association'
22	2	0,3%	filière 'branch', sauce 'sauce'
23	1	0,1%	assurance 'insurance'
25	1	0,1%	coin 'corner'
29	1	0,1%	accès 'access'
35	1	0,1%	rubrique 'rubric'
36	2	0,3%	soirée 'evening', papier 'paper'
40	1	0,1%	service 'service'
42	1	0,1%	atelier 'workshop'
46	1	0,1%	secteur 'sector'
63	1	0,1%	rayon 'department'
81	1	0,1%	pôle 'pole'

Table 6: Frequency spectrum of N2s in "new" SUB-NNs (FrWac)

f	V(f)	V(f) rel.	Examples of N2s
1	431	61,7%	
2	108	15,5%	
3	55	7,9%	
4	21	3,0%	
5	19	2,7%	
6	11	1,6%	
7	6	0,9%	
8	5	0,7%	
9	4	0,6%	
10	3	0,4%	
11	2	0,3%	
12	2	0,3%	
13	1	0,1%	
14	2	0,3%	
15	2	0,3%	
16	4	0,6%	
17	1	0,1%	
18	2	0,3%	
19	3	0,4%	cadeau 'gift', enfant 'child', télé 'tv'
20	1	0,1%	internet 'internet'
22	1	0,1%	cinéma 'cinema'
23	1	0,1%	aluminium 'aluminium'
26	2	0,3%	photo 'photo' radar 'radar'
30	1	0,1%	beauté 'beauty'
32	1	0,1%	bidon 'fake' (lit. 'can')
34	1	0,1%	zen 'calm' (lit. 'zen')
35	1	0,1%	santé 'health'
37	1	0,1%	bébé 'baby'
38	2	0,3%	auto 'car', maison 'home-made' (lit. 'house, home')
41	1	0,1%	achat 'purchase'
42	1	0,1%	papier 'paper'
48	1	0,1%	jeunesse 'youth'
58	1	0,1%	client 'customer'
65	1	0,1%	radio 'radio'