



Are French NNs variants of N-PREP-N constructions? A corpus-based study of two competing patterns

Jan Radimský (University of South Bohemia in České Budějovice)

ABSTRACT

This study aims to provide a thorough empirical examination of the hypothesis that French subordinate Noun-Noun compounds (*stylo-bille* — ‘ballpoint pen’) are mere variants of corresponding phrasal lexemes or syntactic phrases with the structure Noun-Prep-(Det)-Noun (*stylo à bille*). On the basis of extensive corpus data from FrWac, it will be argued that the relationship and the competition between French NNs and NPNs differ with respect to different subtypes of NNs. On the one hand, attributive NNs cannot have NPN variants, and appositive NNs as well as NNs in which the N₂ has a bound meaning have synonymous NPN variants only occasionally. On the other hand, for subordinate verbal-nexus NNs the NPNs represent stylistic variants which seem to be always available. The case of subordinate ground NNs proves to be more complex since, in this case, both patterns are in competition (Aronoff 2016); data discussed in this paper indicate that this competition is steered by phenomena of constructionalization. Since French NNs tend to be organized around paradigmatic families with repeated components, we have put forward the hypothesis that such paradigmatic regularity underpins a progressive formation of ‘niches’ in which new subordinate ground NNs win the competition with NPNs. Moreover, mechanisms of constructionalization even give rise to new subpatterns of subordinate ground NNs whose NPN variants are ungrammatical. The competition between French NNs and NPNs, also documented on diachronic data from Google n-grams, reflects a change in naming strategies in French, especially from the 1960s onwards.

KEYWORDS

competition, compounding, construction morphology, constructionalization, French, niche, phrasal lexeme, word formation

RÉSUMÉ

Les NN français sont-ils des variantes des constructions N-PREP-N ?

Étude de deux modèles concurrents basée sur corpus

Cet article vise à examiner empiriquement l’hypothèse selon laquelle les composés français de subordination du type Nom-Nom (*stylo-bille*) sont de simples variantes des synapsies ou des syntagmes nominaux correspondants avec la structure Nom-Prép-(Det)-Nom (*stylo à bille*). En nous appuyant sur de nombreuses données extraites du corpus FrWac, nous défendons l’hypothèse selon laquelle la relation et la concurrence entre les NN et les NPN français diffèrent en fonction des sous-types de NN. D’une part, les NN attributifs ne peuvent pas avoir de variantes NPN, et les NN appositifs ainsi que les NN dans lesquels le N₂ a un sens lié n’ont des variantes NPN synonymes qu’occasionnellement. En revanche, pour les NN subordonnés « verbal-nexus » les NPN représentent des variantes stylistiques qui semblent toujours disponibles. Le cas des NN subordonnés « ground » se révèle plus complexe puisque dans ce cas, les deux schémas sont en compétition (Aronoff 2016); les données discutées dans cet article indiquent que cette compétition est conditionnée par des phénomènes de constructionalisation. Étant donné que les NN français tendent à s’organiser autour de familles paradigmatiques à composantes répétées, nous avons avancé l’hypothèse qu’une telle régularité paradigmatique sous-tend une formation progressive de « niches » dans lesquelles de nouveaux NNs de subordination « ground » gagnent sur les NPN. De plus, les mécanismes de constructionalisation donnent même naissance à de nouveaux sous-

patrons des NN pour lesquels les variantes NPN sont agrammaticales. Enfin, grâce aux données diachroniques de Google n-grams, il apparaît que la concurrence entre les NNs et les NPNs reflète un certain changement dans les stratégies de dénomination en français, notamment à partir des années 1960.

MOTS-CLÉS

composé, concurrence, formation de mots, français, morphologie constructionnelle, niche, synapsie

DOI

<https://doi.org/10.14712/18059635.2020.2.4>

1 INTRODUCTION

This paper aims at providing a thorough empirical examination of the hypothesis that French subordinate Noun-Noun compounds, such as *roman* [*photos*]_N ('photo novel'), are mere variants of corresponding syntactic phrases or phrasal lexemes, such as *roman* [*avec (des) photos*]_{pp} (cf. Fradin 2009: 433, amongst others).

It is a notoriously known fact that in modern Romance languages, the pattern structured as N-PREP-(DET)-N (e.g. *chemin de fer* — 'railway') is the default and very productive way to express the subordinate relationship between two nouns. As far as French is concerned, already the seminal analysis of French N-PREP-N constructions (henceforth 'NPNs') by Benveniste (1974: 171–176) showed that these constructions, labelled as 'synapsies' by this author, play a crucial role in the process of naming, even though their formal shape resembles rather that of syntactic phrases. To put it differently, French NPNs are commonly used with reference to concepts that, in languages from other families, tend to be expressed rather by other formal means, such as NN compounds (e.g. in Germanic languages) or NA/AN phrases with denominal adjectives (as in Slavonic languages).¹

Although the Noun-Noun pattern also exists in modern French, for a long time it has been reserved to express the attributive relationship between nouns only, i.e. for cases, such as *bourgeois gentilhomme* ('bourgeois gentleman', cf. Rainer and Buridant 2015). Nevertheless, there is evidence that sparse cases of subordinate Noun-Noun compounds have begun to appear in French at least since the beginning of the 19th century (cf. Darmesteter 1874; Hatcher 1946, Arnaud 2003; Villouing 2012 or Rainer and Buridant 2015, amongst others) and empirical corpus data analyzed in Radimský (2019) show that an exponential growth of both type and token frequency of subordinate NNs has been observed in French since the 1960s. Consequently, the new French subordinate NNs are in direct competition with the well established pattern of NPNs, in the sense of Lindsay and Aronoff (2013) or Aronoff (2016),² and it thus seems reasonable to ask what this competition looks like.

1 Cf., for instance, the French NPN *fil d'acier* (wire PREP steel), the English NN *steel wire* and the Czech AN *ocelový drát* ([[steel]_N.SUFF]_A wire). Notice that French has also a vital NA pattern with denominal adjectives; its competition with the NPN pattern has been empirically explored by Jana Strnadová (2014: 93–100).

2 Cf. Gardani et al. (2019) for an overview.



For the time being, there is no agreement on the question in the literature. On the one hand, Fradin (2009) claims that French subordinate NNs (referred to as ‘two-slot nominal constructs’ by the author) always have an NPN counterpart, which entails that subordinate NNs would be mere stylistic variants of the corresponding NPNs. On the other hand, Arnaud (2015) argues that for many subordinate NNs the corresponding NPN construction is either attested in negligible numbers only or is even impossible to form: in such a case, it could be assumed that at least some subordinate NNs fulfill some specific function that differs from the function of NPNs. As Arnaud (2015) suggests, the question behind this controversy is twofold. First, one may ask whether any French subordinate NN can in principle have an NPN counterpart. Second, in the cases where this possibility exists, the question is whether the attested subordinate NNs really have NPN variants in actual utterances and if not, for what reason. In other terms, the first question concerns the potentiality within language system (the Saussurean ‘langue’), while the second one concerns the use of language (‘parole’).

This paper is structured as follows: Section 2 will introduce basic theoretical assumptions comprising a detailed classification of French NNs (2.1.), a brief discussion of French NPNs (2.2.) and an outline of preliminary hypotheses concerning the question which types of NNs are expected to have synonymous NPN variants (2.3.). Section 3 will present an empirical analysis based on 3,370 French NNs (types) and the corresponding NPNs gathered from the corpus FrWac, in order to empirically verify the hypotheses discussed in Section 2.3. Then, Section 4 will outline some emerging ‘niches’ for the subordinate NN pattern that derive from the process of constructionalization (in the sense of Traugott and Trousdale 2013). Additional data from Google n-grams (henceforth GN) will be used in order to visualize the competition between selected French subordinate NNs and NPNs in diachrony. From the theoretical and terminological points of view, the paper is rooted within the framework of Construction Morphology (henceforth CM; Booij 2010; 2016) and the Scalise-Bisetto (2009) classification of compounds, though the empirical data analysis will be relevant to any theoretical framework.³

On this basis, it will be argued that the competition between French subordinate NN compounds and the corresponding NPN phrasal lexemes or noun phrases is a complex phenomenon that documents a change in naming strategies in French, especially from the 1960s onwards.

2 PRELIMINARIES

2.1 TYPOLOGY OF FRENCH NNS

In compliance with the Scalise-Bisetto (2009) general classification of compounds, contemporary French displays three basic types of left-headed endocentric Noun-

3 In this paper, we will adopt a unitary approach assuming that all French NNs may be considered as ‘compounds’. There are alternative views in the literature (for a recent overview see Van Goethem and Amiot 2018), but a different stance with regard to this question has only a limited impact on the issue discussed here.



Noun compounds, namely coordinate ('COORD') NNs, attributive-appositive ('ATAP') NNs and subordinate ('SUB') NNs. Each of the latter two groups may be subsequently split into two subtypes, as indicated in Table 1.⁴ This section will outline key properties of the different subtypes with regard to French Noun-Noun compounds and in Section 2.3., preliminary hypotheses about their potentiality to have synonymous NPN variants will be discussed.⁵

Compounds				
Coordinate (COORD) (or Reversible attributive)	Attributive-Appositive (ATAP)		Subordinate (SUB)	
	Attributive	Appositive	Verbal-nexus	Ground
<i>chanteur-compositeur</i> 'singer-composer'	<i>bourgeois-gentilhomme</i> 'bourgeois gentleman'	<i>oiseau-mouche</i> 'hummingbird' lit. 'bird-fly'	<i>exposition photos</i> 'photography exhibition'	<i>roman photos</i> 'photo novel'

TABLE 1. French endocentric Noun+Noun compounds

Both Attributive-Appositive ('ATAP') NNs and Coordinate ('COORD') endocentric NNs feature a head-attribute relationship between the components, which entails that such a relationship is paraphrasable by an attributive 'is' paraphrasis. However, the exact shape of such paraphrasis is slightly different for each subtype, as indicated in the respective examples (1a-c) with reference to the formalism of the Construction Morphology framework.

- (1a) Attributive NN compound
bourgeois gentilhomme — 'bourgeois gentleman'
 $[N_i N_j]_{Nk} \leftrightarrow [SEM_i \text{ is a } SEM_j]_k$

4 French Noun-noun compounds are almost exclusively endocentric barring a handful of exocentric coordinate dvandva compounds with mostly additive interpretation, such as *physique-chimie* ('physics-chemistry', as teaching disciplines). As far as the position of the head is concerned, most of French NNs are left-headed. Some apparently right-headed attributive compounds, such as *bébé-phoque* ('baby seal', cf. Villoing 2012: 39), as well as right-headed subordinate N-N compounds that have origins in neoclassical compounding pattern, such as *autoroute* ('highway', lit. 'car+way'), will be disregarded in this analysis, since they do not enter in competition with NPN structures.

5 The Scalise-Bisetto typology of compounds has already been discussed thoroughly in relation to Italian NNs by Radimský (2015: 85-160), and it turns out that most of the arguments are also relevant for French NNs. The following paragraphs will therefore provide only the gist of the discussion concerning the general Scalise-Bisetto framework and its applicability on Romance NNs, and readers will be referred to Radimský (2015) for details.



(1b) Coordinate NN compound (= Reversible attributive NN)

chanteur-compositeur — ‘singer-composer’
$$[N_i N_j]_{Nk} \leftrightarrow [SEM_i \text{ is a } SEM_j]_k \ \& \ [SEM_j \text{ is a } SEM_i]_k$$
compositeur-chanteur — ‘composer-singer’
$$[N_i N_j]_{Nk} \leftrightarrow [SEM_i \text{ is a } SEM_j]_k \ \& \ [SEM_j \text{ is a } SEM_i]_k$$

Attributive (1a) and Coordinate (1b) NNs feature a genuine head-attribute relationship between N1 and N2, respectively. The peculiarity of Coordinate NNs (1b) with respect to Attributive NNs consists in the fact that in the former type, the concepts expressed by the compound components are semantically related and situated on the same level of conceptual hierarchy (for instance in (1b), both N1 and N2 represent two names of professions), which means that the order of constituents may be inverted. In the case of component inversion, coordinate compounds remain well acceptable and maintain — more or less — the same meaning (compare the two versions in 1b), while attributive compounds become either ungrammatical or acquire a very different meaning.⁶ Therefore, Coordinate NNs may also be referred to as Reversible attributive NNs (Radimský 2015: 158).

In the subclass of appositive compounds (1c), the attributive relationship between the components is less straightforward, which makes this subclass closer to the Subordinate group. In the original proposal by Scalise and Bisetto (2009), appositive compounds lack clear delimitation: they were defined broadly as compounds in which the modifier is expressed by a noun that “often” has a metaphoric interpretation (Scalise and Bisetto 2009: 51–52). Following the discussion in Radimský (2015: 93–102) and with reference to similar arguments outlined for French NNs by Arnaud (2003: 12–13) or Fradin (2009: 430–431), appositive NNs will be defined here as those in which the attributive ‘is a’ relationship requires an additional operator in the paraphrasis, such as ‘is like a’, because only selected semantic properties of the modifier (N2) apply on the head.⁷ Examples of the respective paraphrases using the CM formalism are given in (1c).

(1c) Appositive NN compound

oiseau-mouche — ‘hummingbird’, lit. *bird-fly*

$$[N_i N_j]_{Nk} \leftrightarrow [SEM_i \text{ is like a } SEM_j]_k$$

(‘a bird that is like a fly’)

6 For a thorough discussion of this issue, see Radimský (2015: 102–112), cf. also Fradin (2009: 430).

7 Arnaud (2003: 12–13) makes a distinction between “total categorial identity” (Fr. *identité catégorielle totale*) and the “partial identity” with reference to the semantic relationship between the compound components. Also notice that the use of similar arguments does not necessarily entail similar terminology: “appositive compounds”, as defined in this paragraph, correspond in fact to “subordinate compounds” in Fradin (2009) and to “attributive compounds” in Villoing (2012).

- $[N_i N_j]_{Nk} \leftrightarrow [SEM_i \text{ is a } SEM_j \text{ from the point of view of } X]_k$ (cf. Fradin 2009: 431)
 ('a bird that is like a fly from the point of view of its size')
 $[N_i N_j]_{Nk} \leftrightarrow [SEM_i \text{ resembles } SEM_j]_k$ (cf. Arnaud 2003: 12–13)
 ('a bird that resembles a fly')



Subordinate compounds were defined as those in which the two components share a head-complement relation (Scalise and Bisetto 2009: 45). Following Scalise and Bisetto (2009: 50–51), the group of subordinate NNs may also be divided into two subtypes: Verbal-nexus NNs and Ground NNs, as indicated in (2a) and (2b), respectively. Verbal-nexus NNs feature either a verb-argument or verb-complement/adjunct relationship between the deverbal head and non-head component, while the components of Ground NNs may have any other type of subordinate relationship.

- (2a) Subordinate Verbal-nexus N+N compound
exposition photos — ‘photography exhibition’
 $[V X(\text{suff.})_b]_{Ni} \leftrightarrow [SEM_a SEM_b]_i$
 $\approx [N_i N_j]_{Nk} \leftrightarrow [SEM_i \text{ having } SEM_j \text{ as argument/complement/adjunct}]_k$
- (2b) Subordinate Ground N+N compound
roman-photos — ‘photo novel’
 $[N_i N_j]_{Nk} \leftrightarrow [SEM_i \text{ in } R \text{ with } SEM_j]_k$

The idea of setting apart a subtype of subordinate compounds with a (de)verbal head has a long tradition in morphology.⁸ Suffice it to say that, for the purpose of this analysis, the subclass of Verbal-nexus NN compounds will cover not only NNs with a suffixed head noun (3a), traditionally called “synthetic compounds”, but also those with a converted head noun (3b–c), referred to as “non-affixal deverbal compounds” by Lieber (2016: 45). It is also to be noticed that French verbal-nexus NNs may be interpreted either as result nouns (3a–b) or event nouns (3c), although the data in our sample suggest that the first option is much more common.

- (3a) *directeur marketing* — ‘marketing director’
 (3b) *accès pompiers* — ‘emergency entrance zone for firemen’
 (3c) *sortie nature* — ‘nature outing’⁹

8 It can be traced back to earlier oppositions between root or primary compounds, on the one hand, and synthetic or secondary compounds, on the other. The term “verbal-nexus” is attributed to Allen (1978) by Bauer (2001: 701), while Scalise and Bisetto (2009: 39) trace it back to Marchand (1969). From a more general perspective, however, it refers undoubtedly to the French term “noeud verbal” used by Lucien Tesnière (1959) to describe the core syntactic function of the verb within the sentence.

9 *La prochaine sortie nature aura lieu le dimanche 17 octobre à l’île Callot où nous profiterons de la marée basse pour faire un grand tour de l’île.* (FrWac) — ‘The next nature outing will take place on Sunday October 17 at Callot island where we will take advantage of the low tide to take a grand tour of the island.’



Verbal-nexus NNs still seem to be rare in French to the extent that Villoing (2012:50) has claimed that the process-argument relationship is impossible within French NN compounds. A corpus-based study by Radimský (2018a) revealed that the pattern exists in French, even though it lacks the well-known regularity of Italian Verbal-nexus NNs. In the sample gathered for the purpose of this analysis, Verbal-nexus NNs represent 15% (490 types).

Another issue raised about Romance Verbal-nexus NNs (see, among others, Lieber and Scalise 2006) — and extended by Fradin (2009: 432–433) to the whole group of French subordinate NNs — concerns their alleged syntactic origin, as they seem to allow for lexical integrity violation. In short,¹⁰ Fradin (2009) claims that all subordinate NNs are not compounds, but rather “two-slot nominal constructs” because they allow for insertion of phrases (4a) and they systematically have a synonymous NPN variant (4b).

- (4a) *alliage* [[*haute*]_A [*température*]_N]_{NP} — ‘high temperature alloy’
 (4b) *alliage* [[à]_P [[*haute*]_A [*température*]_N]_{NP}]_{PP} — ‘high temperature alloy’

The latter property will be commented on in the following sections of this article. As for the former property, Radimský (2015: 230–233) has shown that in the case of Italian NNs, the possibility of phrase insertion does not draw the division line between subordinate NNs, on the one hand, and appositive NNs, on the other, as Fradins suggests.¹¹ Indeed, the difference rather lies in the fact that subordinate NNs allow for (free) insertion of phrases on the non-head component (4a-b), while ATAP (including both attributive and appositive NNs) tend to allow for insertion of phrases on the head component (5a-b). This is explained by the fact that the interpretation of subordinate compounds originates in the head noun, while in ATAP compounds it is triggered by the non-head noun. In other words, only the component that does not trigger the interpretation of the compound is prone to having an internal modifier.

- (5a) [[*valore di concentrazione*]_{NPN} *limite*] — ‘maximum concentration value’ (attributive NN)
 (5b) [[*settore tecnologico*]_{NA} *chiave*] — ‘key technology sector’ (appositive NN, modifier with a bound meaning)

Similar examples of ATAP NN compounds with the head noun slot filled by an NP may also be found in French (6a-b), and they seem well acceptable.¹² Therefore, all types of French NNs outlined in Table 1 will be considered as morphological constructs.

¹⁰ See Radimský (2015: 30–35 and 52–63) for a detailed discussion of the Lexical integrity hypothesis with reference to Italian NNs.

¹¹ Notice that Fradin (2009) operates with a different terminology: he uses the term “two-slot nominal construct” for our subordinate compounds, while his category of “subordinate compounds” applies to compounds labelled as “appositives” in the present analysis.

¹² The examples (6a-b) were identified in FrWac and have thousands of hits in a Google search as “exact match”.

- (6a) $[[\textit{angle de réfraction}]_{\text{NPN}} \textit{limite}]$ — ‘maximum refraction angle’ (attributive NN)
 (6b) $[[\textit{secteurs économiques}]_{\text{NA}} \textit{clés}]$ — ‘key economy sectors’ (appositive NN, modifier with a bound meaning)



This overview of the typology of French NNs is important because the terms discussed above are not used consistently across the different analyses of French NNs in the literature, even when they explicitly refer to the Scalise-Bisetto classification. For instance, Villoing (2012) refers to appositive compounds in (1c) as “attributives”. Fradin (2009) calls attributive compounds in (1a) “subordinate” and subordinate compounds in (2a-b) are referred to as “two-slot nominal constructs”. In this paper, the terminology introduced in Table 1 is consistently used even when referring to secondary sources with a different terminological background.

However, the consistency of the typology outlined in this section is challenged by NN compounds in which one of the components (usually the non-head noun) systematically has a specific bound meaning in the NN pattern. Considered in isolation, some of these compounds allow for the ‘is like a’/‘resembles’ paraphrasis which would make it possible to consider them as appositives (7a-b), while others do not allow for such paraphrasis (7c-7e), so that they should rather be classified as subordinate NNs.

- (7a) *mot clé* — ‘keyword’, ‘very important word’
 $[N_i \textit{clé}]_{\text{Nk}} \leftrightarrow [\textit{very important SEM}_i]_{\text{k}}$
 Appositive paraphrasis: ‘a word that is like a key’
- (7b) *ambiance zen* — ‘zen atmosphere’
 $[N_i \textit{zen}]_{\text{Nk}} \leftrightarrow [\textit{‘calm, peaceful’ SEM}_i]_{\text{k}}$
 Appositive paraphrasis: ‘an atmosphere that resembles zen’
- (7c) *confiture maison* — ‘homemade jam’
 $[N_i \textit{maison}]_{\text{Nk}} \leftrightarrow [(\textit{meal}) \textit{SEM}_i \textit{‘made at home’}]_{\text{k}}$
- (7d) *yaourt nature* (= $[au]_{\text{DetPrep}} [\textit{nature}]_{\text{A/N?}}$) — ‘plain yoghurt’ (nature — metonymic)
 Possible NPN variant: $[yaourt]_{\text{N}} [au]_{\text{DetPrep}} [\textit{nature}]_{\text{A/N}}$
 $[N_i \textit{nature}]_{\text{Nk}} \leftrightarrow [(\textit{meal}) \textit{SEM}_i \textit{‘without added flavouring’ and/or ‘organic’}]_{\text{k}}$
couscous nature — ‘plain couscous’
chocolat nature — ‘plain chocolate’ or ‘organic chocolate’
- (7e) *excuse bidon* — ‘lame excuse’
 $[N_i \textit{bidon}]_{\text{Nk}} \leftrightarrow [\textit{‘lame/phony’ SEM}_i]_{\text{k}}$
argument bidon — ‘phony argument’
annonce bidon — ‘phony ad’

However, the bound interpretation of the non-head noun (N₂), based usually on its metaphorical or metonymical interpretation, brings about the possibility of a genuine attributive ‘is a’ paraphrasis which seems conceivable for most of these com-



pounds, at least in contemporary colloquial French. Indeed, the strings ‘*ambiance est zen*’, ‘*confiture est maison*’, ‘*yaourt est nature*’, ‘*excuse est bidon*’ and ‘*innovation est clé*’ (‘innovation is important’) have between several hundreds to several thousands hits in Google search (as exact string, i.e. in quotation marks).¹³ Following this argument, the compounds in (7a–e) would have to be considered as attributives.

To sum up, French NN compounds in which the non-head component has a specific bound meaning within the NN pattern pose a serious challenge to the classification, so that the decision whether to put them in the attributive, the appositive, or the subordinate group will be arbitrary to some extent. Therefore, these compounds will be considered as a specific category in our analysis.

2.2 TYPES OF FRENCH NPNS

From a purely formal point of view, two structures are in theory possible for French NPNS, in relation to whether the second (i.e. the non-head) noun is introduced by a bare prepositional phrase (8a) or by a prepositional phrase accompanied by a determiner (8b). Though the former case is more typical for phrasal lexemes and the latter for (free) syntactic phrases, this correlation is too blurry to be used as a formal criterion, as will be briefly demonstrated in this section.

(8a) *code à barres* — ‘bar code’ (phrasal lexeme, no determiner)
N-PREP-N

(8b) *code de la route* — ‘highway code’ (phrasal lexeme, definite article)
N-PREP-DET-N

From the semantic perspective, non-head elements in phrasal lexemes typically have generic reference (9a), while those in syntactic phrases have specific reference (9b).

(9a) *directeur d’usine* — ‘factory manager’ (phrasal lexeme, no determiner, generic reference)

(9b) *directeur de l’usine* — ‘manager of the factory’ (syntactic phrase, definite article, specific reference)
(Strnadová 2008: 46)

13 Some examples:

Ad (7a): *Or, leur rôle est clé.* — ‘Well, their **role is key** [=important].’ (FrWac)

Ad (7b): *Les serveurs sont très gentil, l’ambiance est zen, je le recommande!* — ‘The servers are very nice, the atmosphere is calm, I recommend it!’ (FrWac, the missing final ‘s’ in *gentil* reproduces the typo in the corpus.)

Ad (7c): *La purée est maison, les frites aussi.* — ‘The mash is homemade, the fries too.’ (FrWac)

Ad (7d): *Le yaourt est nature c’est à dire sans sucre ni fruits.* — ‘The yogurt is natural, i.e. without sugar or fruit’ (Google search).

Ad (7e): *Ton excuse est bidon!* — ‘Your excuse is bogus!’ (FrWac)



However, this semantic opposition correlates only partly with the formal presence or absence of the determiner behind the non-head noun in the way it can be observed in examples (9a) and (9b). On the one hand, as Strnadová (2008: 45–50) points out, absence of the determiner is usually well correlated with the generic reference, except for some particular cases, such as when the non-head noun denotes days of the week (10a–b).¹⁴ On the other hand, presence of the determiner does not entail specific reference: it is far from exceptional to encounter cases when the non-head noun in French NPNs has generic reference despite of being preceded by a definite determiner (8b, 11b).

(10a) *journal de samedi* — ‘last Saturday’s newspaper’ (specific reference, no determiner)

(10b) *journal du samedi* — ‘Saturday newspaper’ (generic reference, determiner)
(Strnadová 2008: 49)

(11a) *fête de famille* — ‘family party’ (generic reference, no determiner)

(11b) *fête de la famille* — ‘family day’ (generic reference, definite determiner, but different meaning with respect to 11a)
(Strnadová 2008: 46)

To put it differently, NPNs containing bare prepositional phrases, such as (8a, 9a), are quite likely to be phrasal lexemes and thus potentially synonymous equivalents of NNs.¹⁵ However, the same also holds true for many NPNs containing prepositional phrases with a definite article, such as (8b, 10b, 11b). So technically speaking, all NPNs will be taken into account in the data, irrespective of the fact whether the non-head noun has an overtly expressed article or not.

2.3 HYPOTHESES ABOUT EXPECTED NPN VARIANTS OF NNS

This section will provide an overview of preliminary hypotheses concerning the question whether the different types of French NNs introduced in 2.1. are expected to have (synonymous) NPN variants. It will be argued that the expectations are rather clear for subordinate verbal-nexus NNs and for attributive NNs, while hypotheses concerning the other types of NNs seem less reliable.

14 Not to mention that according to the French grammar, the preposition *de* combined with the plural form of the indefinite article (*des*) yields the bare form *de*, homonymous with the bare preposition. Therefore, a sequence such as *transportN deP+Art.Ind. matièresN.pl. dangereusesA* (‘transport of hazardous materials’), which on the surface seems to contain an NPN with a bare PP (*transport [deP matièresN.pl.]_{PP}*), is the plural form of *transportN d’P uneArt.Ind. matièreN.pl. dangereuseA* (‘transport of a hazardous material’). Such homonymy is challenging for computational treatment of data.

15 In Romance NN compounds made up of two common nouns, the non-head noun can almost always have only generic reference (cf. also Haspelmath and Sims 2010: 191). It seems that the possibility to have an argument with a specific reference is a peculiarity of some Italian Verbal-nexus NNs, as is argued in Baroni, Guevara and Zamparelli (2009: 33).



Subordinate verbal-nexus NNs — or at least those in which the argument corresponds either to the subject or to the direct object of the verb underlying the deverbal head — are always supposed to have a possible NPN paraphrasis with the preposition *de*, as illustrated in (12a-c). Indeed, French deverbal nouns, irrespective of whether they bear eventive or resultative meaning, usually allow for expressing arguments (or at least one argument at one time) in a prepositional phrase introduced by *de* (Samvellian 1995: 106–188). Moreover, an empirical study by Baroni, Guevara and Pirrelli (2009) shows that more than 90% of Italian verbal-nexus NNs (types) indeed have an NPN counterpart with the preposition *di* between the same pair of nouns in the same corpus. Similar empirical results could therefore be expected for French.

- (12a) *exposition photos* — ‘photography exhibition’
 (12b) *exposition de photos* — ‘photography exhibition’
 (12c) *exposition des photos* — ‘photography exhibition’

Conversely, attributive NNs in theory should not have NPN variants because the attributive relation, paraphrasable by “IS A” (13a), can hardly be reformulated by an NPN phrase (13b). Indeed, such possibility exists in French only for exceptional subschemas, namely when the head noun denotes a month (*mois de mai* — ‘May’, lit. ‘month of May’), a toponym (*ville de Paris* — ‘city of Paris’) or an insult (*ce crétin de Paul* — ‘this moron Paul’, lit. ‘this moron of Paul’). The same arguments should also apply to coordinate endocentric NNs which, as has been demonstrated in 2.1., represent a specific reversible variant of attributive NNs.

- (13a) *(le) pays membre* — *(le) pays EST (UN) membre*
 ‘(the) member country’ — ‘(the) country IS (A) member’
 (13b) *(le) pays *PREP membre*
 ‘(the) country *PREP member’

The case of appositive NNs seems similar to that of attributive NNs. On the one hand, the paraphrases for appositive compounds (14a) are quite similar to those of attributive NNs (13a) and Fradin (2009: 430–433) also claims that appositive NNs, such as (14a), cannot have synonymous NPN variants. On the other hand, it is not impossible to conceive that PPs with certain prepositions (especially *à* or *en*) could express partial resemblance in aspects, such as ‘dimension’ or ‘shape’ between the concepts referred to by N₁ and N₂ (14c).

- (14a) Appositive NN compound
oiseau-mouche — ‘hummingbird’, lit. *bird-fly*
 $[N_i N_j]_{N_k} \leftrightarrow [SEM_i \text{ is like a } SEM_j]_k$
 (‘a bird that is like a fly from the point of view of its size’)
 $[N_i N_j]_{N_k} \leftrightarrow [SEM_i \text{ resembles } SEM_j]_k$
 (‘a bird that resembles a fly’)



(14b) *oiseau ?PREP mouche*
 ‘bird ?PREP fly_N’

(14c) *verre (à) ballon* — ‘balloon glass’
 (‘a glass that resembles a balloon’)

In other words, appositive NNs should not have synonymous NPN variants systematically, but occasionally they could.

The case of ground subordinate compounds, which are of major concern in this research, is controversial, too. As pointed out in the introduction, there is no agreement in the literature either on the question whether all French ground NNs may have a potential NPN paraphrasis, or on the question whether such paraphrasis — if it sounds acceptable to native speakers — is in real use (cf. Fradin 2009; Arnaud 2015). From the theoretical point of view, it seems reasonable to expect that any GROU-NN should be at least potentially paraphrasable by an NPN — as Fradin (2009) suggests — because, since contemporary French does not express the morphological category of case through nominal inflection, prepositional phrases are the default way to express subordinate relations within noun phrases. However, this hypothesis needs further empirical verification.

As for NN compounds in which the non-head noun has a bound meaning, it seems reasonable to assume that the possibility that such NNs would have synonymous NPN variants is rather unlikely and accidental. This is due to the fact that the difference between the ‘bound’ meaning of the non-head noun within the NN pattern and the ‘non-bound’ meaning of the same noun within the NPN pattern blocks a systematic synonymy between the respective NNs and NPNs (15a-b), provided that both variants are attested at all.

(15a) *système clé* — ‘key system’, ‘important system’

(15b) *système de clés* — ‘system of keys’

Table 2 provides a comprehensive overview of the hypotheses that will be subject to empirical verification in the following section.

Type	Subordinate Verbal-nexus	Attributive (& Coordinate)	Appositive	Subordinate Ground	N2 with bound meaning
Example	<i>exposition photos</i> ‘photography exhibition’	<i>bourgeois-gentilhomme</i> ‘bourgeois gentleman’	<i>oiseau-mouche</i> ‘humming-bird’ lit. bird-fly	<i>roman photos</i> ‘photo novel’	<i>ambiance zen</i> ‘zen atmosphere’
NPN variants expected	yes	no	probably no	probably yes	probably no

TABLE 2. Expected NPN variants of French endocentric Noun+Noun compounds



3 ANALYSIS OF DATA FROM FRWAC

3.1 DATA EXTRACTION

Data for this research were extracted from FrWac, a very large web corpus (1600M tokens) developed between 2005 and 2007, which contains mainly texts written around the year 2000 and pertaining to various genres.

As for the NNs, the extraction of the sample was based on the analyses of complete frequency lists of binominals in either loose (16a) or hyphenated (16b) forms preceded by an article or a preposition,¹⁶ with a minimum token frequency set to 3.¹⁷

(16a) *le centre ville* — ‘town centre’

(16b) *le centre-ville* — ‘town centre’

From the final merged frequency list of binominals with over 430,000 NN types, a sample of 3,370 types (corresponding to more than 600,000 tokens) was filtered out manually. Results of sampling strategies from previous research were used (Radimský 2018a; 2018b; 2020), but compounds were re-checked with regard to the categories outlined in Section 2.1. The proportion of the different categories of NNs in the sample is given in Table 3.

Category		Total of NNs (type fq.)	Rate of NNs	
ATAP	Appositive	186	5.5%	23.4%
	Attributive	604	17.9%	
Subordinate	Subordinate Verbal nexus	490	15.0%	62.0%
	Subordinate Ground	1585	47.0%	
ATAP or Subordinate	Containing N2 with a bound meaning ¹⁸	505	14.5%	14.5%
Total:		3370	100%	

TABLE 3. Overview of type frequencies of NNs for each category in the sample

¹⁶ The presence of an article or a preposition helps eliminating undesired noise introduced by long sequences of nouns.

¹⁷ Since FrWac is a web corpus in which texts have been gathered automatically without manual cleaning, the items with $fq. < 2$ contain too much noise to allow for a feasible procedure of compound filtering. On the other hand, the size of the corpus is such that even items with $fq.$ between 3–4 can be considered as hapax legomena.

¹⁸ The following N2s with a bound meaning have been identified in the NNs from our sample (ordered by decreasing type frequency): *clé* (‘important’), *fantôme* (‘not real’ or ‘abandoned’), *zen* (‘calm, peaceful’), *bidon* (‘lame/phony’), *maison* (‘home-made’), *culte* (‘fanatically admired’), *choc* (‘causing emotional shock’), *éclair* (‘extremely quick’), *étalon* (‘standard’), *relais* (‘intermediary’), *design* (‘designer, expensive and fashionable’), *nature* (‘without added flavouring’ and/or organic’), *cadre* (‘general, framework’), *pilote* (‘trial, pilot’), *minute* (‘very quick’), *fantoché* (‘puppet, non-independent’), *fantaisie* (‘unusual’ or ‘fake’), *galère* (‘exhausting, troublesome’), *bateau* (‘trite’).



The extraction of NPNs proceeded in a very similar manner in order to ensure that figures for NNs and NPNs could be compared. Complete frequency lists of DET-N-PREP-N sequences either with (17a) or without (17b) a determiner introducing the second noun were extracted — the 10 most frequent prepositions, listed in (18), were taken into account — and the same token frequency filter ($fq.>3$) was applied.

- (17a) *le centre de ville* — ‘town centre’
 (17b) *le centre de la ville* — ‘centre of the town’

- (18) Prepositions (= 10 most frequent): *de/en/à/par/sans/sur/entre/pour/avec/sous*

Finally, the NPN sample was compared to the NN sample, so that the lemmatized N1–N2 combinations that were not present in the sample of NNs were filtered out from the sample of NPNs.

3.2 RESULTS

Overall results concerning the number and especially the rate of NNs (lemmatized types) that do not have an NPN variant in the corpus are given in Table 4. The last two columns provide separate figures for NPNs with bare modifiers and for those with modifiers preceded by a determiner, but since these rates are almost identical, we will comment only on global figures given in the first column (rates in bold).

Type	Total of NNs	No NPN (any variant)		No NPN (BarePrep only)		No NPN (ArtPrep only)	
Appositive	186	125	67%	140	75%	162	87%
Attributive	604	431	71%	477	79%	504	83%
Subordinate Verbal nexus	490	44	9%	139	28%	142	29%
Subordinate Ground	1585	723	46%	919	58%	1073	68%
N2 with a bound meaning	505	416	82%	455	90%	455	90%
Total	3370	1739	52%	2130	63%	2336	69%

TABLE 4. Number and rate of NNs (lemmatized types) that do not have an NPN variant in FrWac

When analyzing figures from Table 4, one must bear in mind two important things. First, absence of an NPN variant for a given NN in the corpus does not entail that such a variant is ungrammatical (however, there might be good reasons why speakers do not use it). Second, automatically detected presence of an NPN variant for a given NN in the corpus does not guarantee that the respective NN and NPN are synonymous. Indeed, there might be various reasons why this is not the case, such as the bound meaning of a component in the NN pattern ((19a) vs. (19b)), the polysemy of a component which is preferentially linked to only one pattern (in (20a) *rayon* refers to a ‘section of a department store’, while in (20b) it denotes a ‘(book)shelf’, i.e. by metonymy



a thematic section in a bookstore), the presence of a phrasal lexeme (e.g. *clé publique* in (19c)), and so on.

- (19a) *système clé* — ‘key system, very important system’ (bound meaning of *clé*)
- (19b) *système de clés* — ‘system of keys’
- (19c) *système à [clé publique]* — ‘public key system’ (‘public key’ being a term from cryptography)

- (20a) *rayon librairie* — ‘bookstore department’ (specialized section of a department store)
- (20b) *rayon de la librairie* — ‘bookshelf’ (specialized section of a bookstore)

This issue could be fixed only through a careful manual filtering which is very hard to achieve with such a large amount of data (thousands of types, hundred thousands of tokens). Bearing in mind these precautions, let us take an overview of the results for the different types of NNs.

In the case of **subordinate verbal-nexus** compounds, the results match our expectations very well. Verbal-nexus compounds were always supposed to have a possible NPN paraphrasis and, indeed, this is exactly what we observe for French data: only 9% of those identified in the sample do not have an NPN variant in the same corpus. This means that not only do verbal-nexus NNs have NPN variants, but also that these variants are widely used. A cursory check of the data reveals that the respective NNs and NPNs are synonymous.

Attributive NNs were expected not to have NPN variants and again, the data match this initial hypothesis well: 71% of types do not have an NPN variant at all, while in the case of remaining 29% of types the NPN variants do not appear to be synonymous with the respective NNs. Some prominent examples of non-synonymy are given in (21–23).

- (21a) *ordinateur destinataire* — ‘recipient computer’
- (21b) *ordinateur du destinataire* — ‘recipient’s computer’

- (22a) *réseau partenaire* — ‘partner network’
- (22b) *réseau de partenaires* — ‘network of partners’

- (23a) *famille victime* — ‘victim family’
- (23b) *famille de la victime* — ‘family of the victim’

As for **positive** NNs, it has been argued that they should not have synonymous NPN variants systematically, but occasional NPN variants were not excluded. This is again what the corpus data show. For 67% of types, the corpus does not provide attested NPN variants, which is almost the same figure as in the case of attributive NNs. As far as the remaining two thirds of types are concerned (61 types in absolute numbers), there are cases of non-synonymous NNs and NPNs (24–25), but also cases of synonymous variants, such as (26a–d).



- (24a) *enfant roi* — ‘overindulged child, spoiled child’
 (24b) *enfant du roi* — ‘child of the king’

- (25a) *élément moteur* — ‘driving element’
 (25b) *élément du moteur* — ‘element of the engine’

- (26a) *rôle (de) pivot* — ‘pivotal role’
 (26b) *livre (en) accordéon* — ‘concertina book’
 (26c) *verre (à) ballon* — ‘balloon glass’
 (26d) *dossier (en) béton* — ‘well prepared application’

Similar hypotheses have been put forward concerning NN compounds in which the **non-head noun has a bound meaning**: it has been argued that the possibility for such NNs to have synonymous NPN variants is rather unlikely. Corpus data show that this is the category with the lowest rate of NPN variants, as 82% of types do not display a corresponding NPN in the corpus. In the remaining 18% of types roughly all NNs have a meaning that differs from the meaning of the respective NPN, as illustrated in (27–28).

- (27a) *problème clé* — ‘important problem’
 (27b) *problème de clé* — ‘problem with key’

- (28a) *lieu culte* — ‘iconic place’
 (28b) *lieu de culte* — ‘place of worship’

It is not excluded, however, that the two meanings could be rather close by accident, such as in (29a–b).

- (29a) *ville fantôme* — ‘ghost town’
 (29b) *ville de fantômes* — ‘town of ghosts’

The case of **ground subordinate compounds** is more delicate. As pointed out in Section 2.3, we might expect them to be at least potentially paraphrasable by NPNs, which is the view also shared by Fradin (2009) but challenged by some counterexamples of Arnaud (2015). According to the global figures in Table 4, ground NNs have 46% of types without attested NPN variants, which represents an intermediate position between verbal nexus NNs (with only 9% of types without attested synonymous NPNs), on the one hand, and all remaining categories (with 67–82% of types without NPN variants, most of the NPNs being not synonymous), on the other. A closer look at the data reveals a set of pairs where both the NN and the respective NPN are in use and can be considered as synonymous variants, such as (30a–f), but surprisingly, such cases do not represent the majority: for 46% of types no NPNs are attested and for many others the NPN variants have very low frequency.¹⁹

¹⁹ If token frequency is taken into account, in only 16% of types (249 types in absolute numbers) the token frequency of the NPN is higher than the token frequency of the corresponding NN.



- (30a) *code (à) barres* — ‘bar code’
 (30b) *écran (de) télé* — ‘TV screen’
 (30c) *émission (de) radio* — ‘radio programme’
 (30d) *emplacement (de) camping* — ‘campsite, camping pitch’
 (30e) *espace (de) détente* — ‘lounge’
 (30f) *littérature (de) jeunesse* — ‘youth adult literature’

How to explain that NPN variants of subordinate ground NNs are so underrepresented? From a qualitative point of view, the data make it possible to argue that the absence of NPNs is not due to the fact that the NPN variants would be impossible to make up or would be ungrammatical. Indeed, it is hard to find examples, such as (31–33), where no preposition seems suitable (31) or the NPN variant has a different meaning than the corresponding NN (32a–33b).

- (31) *portrait (?PREP) robot* — ‘facial composite’
 Possible paraphrasis: *portrait/identikit as if it were made by a robot*
- (32a) *coin cuisine* — ‘kitchenette’
 Possible paraphrasis: *area (in a room) used as a kitchen*
 $[coin_i N_j]_{N_k} \leftrightarrow [‘area in a room used as/for/by’ SEM_j]_k^{20}$
- (32b) *coin de la cuisine* — ‘corner of the kitchen’
- (33a) *rayon librairie* — ‘bookstore department’ (specialized section of a department store)
- (33b) *rayon de la librairie* — ‘bookshelf’ (specialized section of a bookstore)

Therefore, the obvious conclusion is that most of NPN variants of ground subordinate NNs could exist, but they are not in use. Leaving aside some evident individual cases of blocking, such as (34a–c), where an established and well lexicalized NN prevents native speakers from using the NPN variant, a very important question arises at this point: what makes speakers resort to the coinage and to the use of NNs, if the traditional NPN pattern seems well suitable for the same purpose?

- (34a) *soutien-gorge* — ‘bra’
 (34b) *zone euro* — ‘eurozone’
 (34c) *bande-annonce* — ‘trailer’

A comprehensive answer to this complex question goes certainly beyond the scope of this paper, but the data collected in our sample make it possible to outline some hypotheses that will be briefly discussed further.

In short, a closer look on the data reveals that a consistent part of NNs in the sample tends to be organized around paradigmatic families with the same N1 or the same N2 component, as illustrated in Table 5. In this concrete case, frequency counts

20 The construction in (32a) even suggests that the N1 *coin* has a bound meaning here, cf. *coin repas* — ‘dining area’, *coin bibliothèque* — ‘library corner’, *coin enfants* — ‘children’s corner’.

tell us that the only attested NPN variant is *pause de midi* ('lunch break'), which is also much more frequent than the respective NN, while for the other binominals the NN variant is the only one attested.



N1+N2	Fq of NNs	Fq of NPNs (any preposition)	Translation
pause+café	594	0	coffee break
pause+cigarette	45	0	smoke break (lit. 'cigarette break')
pause+clope	28	0	smoke break (lit. 'cigarette break')
pause+détente	79	0	relaxation break
pause+fraîcheur	5	0	cool break
pause+goûter	41	0	snack break
pause+lecture	8	0	reading break
pause+midi	6	164	lunch break (lit. 'midday break')
pause+ravitaillement	5	0	refueling break
pause+repas	119	0	lunch break (lit. 'meal break')
pause+sandwich	13	0	sandwich break
pause+thé	29	0	tea break

TABLE 5. Frequency of NNs and NPNs with the head noun *pause* — 'pause, break' in the sample

Interestingly, the peculiarity of *pause de midi* documented by the frequency counts correlates well with diachronic data: according to Google n-grams,²¹ the NPN *pause de midi* has been attested since the 1880s, while the other binominals — if present in GN at all — are attested only in the NN variant and appear for the first time between the 1960s and 1980s.

Progressive conventionalization of regularities based on NN patterns with a lexically filled component, such as the one illustrated in Table 5, have been well documented since the 1960s for other categories of NNs, too. However, in the case of attributive NNs, of appositive NNs and of NNs in which the non-head has a bound meaning, the NPN variant may rarely enter competition with the NN, as illustrated above. In the next section we will discuss the hypothesis that this type of paradigmatic regularity underpins a progressive proliferation of niches in the sense of Lindsay and Aronoff (2013) and Aronoff (2016), which may account for the fact that some French subordinate ground NNs win the competition against the well established NPN pattern. The sample of ground NNs will be analyzed globally, irrespective of whether the NNs have an NPN counterpart or not, in order to find out what these niches may look like.

²¹ <https://books.google.com/ngrams>



4 SPECIFIC NICHEs OF GROUND NNS

It is well known that Romance NNs are more likely to appear in specific linguistic contexts (such as newspaper headlines), in specific genres (typically in ephemera, such as leaflets, price tags, ads or posters) and in texts on determined topics (such as economic and commercial subjects, management or IT).²² These factors may explain different proportions between token frequencies of NNs and NPNs across different specific subcorpora, but they fail to account for the fact that NPN variants of French grounding NNs are so underrepresented or even absent from a general corpus, such as FrWac, as demonstrated in Section 3. In this section, it will be argued that the progressive “division of labour” between French GROU-NNs and NPNs is underpinned by two processes based on constructionalization, as outlined recently in Radimský (2020): the first results from component-based generalizations and the second emerges from further semantic and functional generalizations over paradigms.

Data analyzed in Radimský (2020) show that the spread and conventionalization of the French NN pattern from the 1960s onwards is reinforced by a component-based regularity, i.e. by the family-size effect of repeated components, be it in the left-hand or the right-hand position. That is, new families with either a repeated N₁ or N₂ component — such as *papier* ‘paper’ in (35a-d) and (36a-d), respectively²³ — progressively appear and grow up in the process of constructionalization. In terms of type frequency, the family size effect on N₁ and N₂ appears to be balanced.

- (35a) *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
- (35b) *papier aluminium* (‘aluminium foil’, lit. ‘aluminium paper’): N₂ specifies the material of N₁
- (35c) *papier cuisson* (‘baking paper’): N₂ is an event that N₁ allows to perform
- (35d) *papier cadeau* (‘wrapping paper’, lit. ‘gift paper’): N₂ is an object affected by the event that N₁ allows to perform

- (36a) *impression papier* (‘printing on paper’)
- (36b) *sac papier* (‘paper bag’)
- (36c) *version papier* (‘paper version’)
- (36d) *journal papier* (‘printed newspaper’)
- (36e) *abonnement papier* (‘print subscription’)

²² For Italian, see Dardano (2009: 228) or Baroni, Guevara and Zamparelli (2009). As for Spanish, Moyna (2011: 164) argues that NNs are “frequent in colloquial registers typical of every-day language”. Noailly (1990: 13) suggests that the diffusion of French NNs is favoured by medias.

²³ The fact that the same repeated component (i.e. *papier*) on both positions is exemplified here has no particular importance.



The presence of such families in the lexicon has consequences for diachronic competition between NNs and the respective NPNs (when they exist): if the family of NNs is larger, users may be more likely to choose the NN version than the NPN version in their utterances. This phenomenon can be observed in diachrony on data from Google n-grams,²⁴ even though we have to bear in mind that NNs will be underrepresented in GN with respect to FrWac (indeed, GN is based exclusively on the corpus of Google books where contexts, genres and topics that favour the use of NNs are necessarily less represented with respect to a web corpus, such as FrWac). For instance, graphs in Annexes 1.1.-1.3. that report relative token frequencies of NNs in (35a, b, d) and the corresponding NPNs show that only in the case of (35b) the NPN keeps pace with the corresponding NN, while for (35a) and (35d) the NN version ‘wins’ (GN data are not available for (35c) and the relative NPN).

The nominal components with the highest family size effect, i.e. with the highest type frequency for the N1 and N2 position in subordinate NNs, as identified in Radimský (2020), are given in (37) and (38), respectively.

(37) N1s with the highest type frequency in new subordinate NNs

pôle ‘pole’, rayon ‘department’, secteur ‘sector’, atelier ‘workshop’, service ‘service’, papier ‘paper’, soirée ‘evening’, rubrique ‘rubric’, accès ‘access’, coin ‘section, area of a house or store’, assurance ‘insurance’, filière ‘branch’, sauce ‘sauce’, association ‘association’, guide ‘guide’, espace ‘space’

(38) N2s with the highest type frequency in new subordinate NNs²⁵

radio ‘radio’, client ‘customer’, jeunesse ‘youth’, papier ‘paper’, achat ‘purchase’, auto ‘car’, bébé ‘baby’, santé ‘health’, beauté ‘beauty’, photo ‘photo’, radar ‘radar’, aluminium ‘aluminium’

Notice that technically speaking, the family size effect of repeated components may have effect on the competition between NNs and NPNs irrespective of the fact whether the replication of the same semantic relationship towards the other component is also involved or not. However, things get more complicated when semantics comes into play.²⁶ For instance, the family in (36a-d) is based on a pure component-

²⁴ <https://books.google.com/ngrams>

²⁵ The N2s that appear with a high type frequency in NNs with a bound meaning, such as *maison* ‘home-made’ (lit. ‘house, home’), *zen* ‘calm’ (lit. ‘zen’) and *bidon* ‘fake’ (lit. ‘can’), are not included in this list, because they seldom enter direct competition with NPNs, as shown in Section 3.2.

²⁶ From the perspective of paradigmatic approaches to morphology (for a recent overview, cf. Bonami and Strnadová 2019), the term ‘family’ is used in a rather loose sense in this paper. It refers to (i) compounding patterns based on semi-schematic constructions, i.e. constructions in which one lexeme is specified, such as (39); (ii) compounding patterns based on semi-schematic constructions with a specified semantic relation between the components, such as (41); (iii) compounding patterns as defined in (ii), in which only one sense of the specified lexeme is involved, such as (42). To the best of the author’s knowl-



based regularity without any semantic specification, so that in CM terms it may be captured by the schema (39).

$$(39) [N_i \textit{ papier}]_{Nk} \leftrightarrow [SEM_i \textit{ related to paper}]_k$$

An analysis of the diachronic competition between the respective NNs and NPNs in GN (see Annex 2.1.-2.5.) reveals that NN variants win in binominals (36c-e), but in the case of binominals in (36a-b) NPNs keep winning over NNs (even though the token frequency curve of NNs grows too). This is explained by the fact that NPNs related to NNs (36a-b) are underpinned by strong syntax-based (40a) or sense-based (40b) regularities, while no such regularities exist for NPNs related to (36c-e), as exemplified in (40c-e). In the latter case (36e), for which no NPN variant is attested in GN, the relationship between components is quite complex, so that speakers are not sure about the right preposition: the NPN version *abonnement sur papier* and even the NNN version *abonnement version papier* (40e) are attested as hapax legomena in FrWac.

(40a) *impression sur papier* ('printing on paper')
imprimer qc. sur papier ('to print sth. on paper')

(40b) *sac en papier* ('paper bag')
 $[N_i \textit{ en } N_j]_{Nk} \leftrightarrow [SEM_i \textit{ made of } N_j]_k$

(40c) *version (?sur) papier* ('paper version')

(40d) *journal (?sur) papier* ('printed newspaper')

(40e) *abonnement (?sur / ?version) papier* ('print subscription')

Syntactic paraphrase: *abonnement à la version (sur) papier* ('subscription to the print version')

To sum up, component-based regularities, such as (39), take full effect when they do not have to struggle against syntax- or sense-based regularities of the respective NPNs (40a-b). Moreover, a specific niche of the NN pattern emerges when speakers are not sure which preposition would be suitable to express a very particular type of semantic relationship between components, as exemplified in (36e) and (40e).

As pointed out in the introduction, the process of constructionalization of new subschemas may involve not only component-based regularities analyzed above, but also further semantic generalizations over NN families. Again, this process may start out either from families with a fixed head noun (N1), or from families with a fixed non-head noun (N2).

In the case of semi-schematic constructions with a fixed non-head noun (N2), the semantic generalizations are more likely to be based on a specific fixed mean-

edge, paradigmatic approaches to morphology have not been systematically applied to compounding yet, one of the first attempts is the application on German AN compounds by Gaeta and Angster (2019). Further research is therefore still necessary to clarify the notion of 'family' and its pertinence for the analysis of compounding patterns.

ing of the repeated component, based on a semantic shift of that component. This is a mechanism that gives rise to the NN patterns in which the modifier (N₂) has a bound meaning, as demonstrated in examples (7a-e) above, irrespective of the fact whether the origin of such patterns is based on constructionalization over appositive compounds (with modifiers such as *clé* — ‘key’ → ‘important’ or *zen* — ‘zen’ → ‘peaceful’ in (7a-b)) or on constructionalization over subordinate ground compounds (with modifiers such as *maison* — ‘house’ → ‘home-made’, *nature* — ‘nature’ → ‘organic’ or *bidon* — ‘can, tin’ → ‘lame/phony’ in (7c-e)). However, the semantic shift of the modifier is not a necessary condition in the process of constructionalization — there are modifiers with a literal interpretation, such as *bébé* (‘baby’), which seem to be quite systematically linked with one type of semantic relation towards the head noun (‘for babies’ in (41)). The difference lies only in the fact that when the modifier has a shifted bound meaning in NNs, the NNs based on such semi-schematic construction can hardly have NPN competitors.

- (41) *lit bébé* — ‘baby cot / crib’ (NPN: *lit pour bébé*)
 [N_i *bébé*]_{j,Nk} ↔ [SEM_i for babies]_k
siège bébé — ‘baby car seat’ (NPN: *siège pour bébé*)
vêtement bébé — ‘baby clothing’ (NPN: *vêtement pour bébé*)

The constructionalization of semantic generalizations can also be observed in families with a fixed head noun (=N₁). In this case, the potential niche of the NN pattern consists in its aptitude for capturing conceptual classifications of head nouns. As Fradin (2003:203) points out, a specific function typical for both attributive and subordinate NNs is to denote a subtype/subclass of head nouns N₁; some examples related to ‘spatial classification’ with head nouns *espace+N* (‘area’), *rayon+N* (‘department’) and *coin+N* (‘section, area of a house or store’) were recently analyzed in Koga (2018).²⁷ Data from the corpus FrWac provide extensive support to this hypothesis and make it possible to claim that: (i) NNs may express any type of classification (i.e. not only “spatial” classification); (ii) in the case of polysemous head nouns, the classification is linked to one meaning of that head noun; (iii) within a “conceptual classification”, the NN pattern may capture generalizations that go beyond specific semantic relations between N₁ and N₂, so it can dispense speakers from deciding about the most appropriate preposition; and (iv) the use of NNs offers the advantage to have one and the same formal pattern (NN) for expressing one conceptual classification for a specified N₁.

Let us illustrate these properties with concrete examples. Table 6 gives examples of NNs with the head noun *rayon* that refers to a ‘supermarket department’. Notice

²⁷ Cf. also “classifying subordinatives” in Van Goethem and Amiot (2019). However, we do not agree that for French NNs, being a “classifying subordinate” correlates well with the wordhood and lexical integrity of the item (Van Goethem and Amiot 2019: 141–142). Examples in Table 6 are rather “classifying” than “qualifying” subordinatives, which does not prevent them from allowing for standard phrase insertion in the non-head slot (*rayon [fruits et légumes]* — ‘[fruit and vegetable] department’, *rayon [fruits secs]* — ‘[dried fruit] department’), as demonstrated in Section 2.1.



N1	(PREP)	N2	N2 — perspective of classification
'department' (in a supermarket)	de	<i>glace</i> 'ice cream' <i>fruits</i> 'fruit' <i>livre(s)</i> 'books'	Type of the item
	des	<i>jouets</i> 'toys'	Type of the item (pl.)
	de	<i>beauté</i> 'beauty'	Purpose of the goods
	pour	<i>enfant(s)</i> 'children' <i>femme(s)</i> 'women'	Target customer
	?de	<i>traiteur</i> 'caterer'	Seller (person) / shop
	?(de) (ATTR)	<i>boucherie</i> 'butchery' <i>épicerie</i> 'grocery' <i>librairie</i> 'book store'	Type of the shop

TABLE 6 — NNs capturing the conceptual classification for *rayon* ('supermarket department')

that the construction (42) that underlies the pattern illustrated in Table 6 is linked to only one meaning of the polysemous lexeme *rayon*. In other terms, it is unlikely to encounter *rayon* in the sense of 'ray/beam' as the head noun of an NN,²⁸ which explains the difference in meaning between (43a) and (43b):

(42) [*rayon* N_j]_{Nk} ↔ [department in a store identified by SEM_j]_k

(43a) *rayon de lumière* — 'ray of light'

(43b) *rayon lumières* — 'light department'

The non-head nouns (N2s) capture the conceptual classification of the supermarket departments from quite different perspectives: they may denote types of the items sold in that department, purpose of the goods, target customers, sellers (i.e. persons working in that department) or the relative individual stores. If the same classification were expressed by NPNs, the (polysemous) preposition *de* would seem suitable in many cases, but it does not work everywhere. For the examples with the attributive relationship between the components (i.e. the last row), it does not sound particularly appropriate. In other terms, the use of NNs in the examples from Table 6 makes it possible to capture one conceptual classification by one formal pattern that yields acceptable results irrespective of the semantics of the modifier, which is not the case with all the NPN variants.

That is, cases in which semantic generalizations over NN families concern meanings which fail to be expressed by NPNs represent another systematic way of how specific niches for grounding NNs emerge.

FrWac data show that the same process can in principle be applied on any head noun with any type of conceptual classification (i.e. not only the 'spatial' one): Tables

²⁸ Except for *rayon laser* — 'laser beam', which is explained by the family size effect of the lexeme *laser* used as non-head noun in NNs.

7–10 in Annexes provide additional examples for the head nouns *assurance* ('insurance'), *coin* ('section, area of a house or store'), *pause* ('break'), and *tarif* ('rate, tariff').

The diachronic competition of three couples of NNs and NPNs for the head noun *assurance* ('insurance')²⁹ in GN is reported in Annexes 3.1–3.3. The graphs show an impressive vitality of NN variants — even though an older NPN variant is attested in each case, the frequency of the NN variant grows and wins at some point. As one would expect, NPNs which had previously higher token frequency (see *assurance sur la vie* — 'life insurance') resist for longer.

4 CONCLUSION

The data analyzed in this paper make it possible to argue that the relationship and the competition between French NNs and NPNs differ with respect to different subtypes of NNs. Five categories of NNs have been considered: subordinate verbal-nexus, attributive, appositive, subordinate ground and NNs in which the non-head has a bound meaning.

For four of these five types, the corpus data confirm the initial theoretical assumptions. In the case of subordinate verbal-nexus NNs the NPN variant was expected to be always available and the corpus data confirm the presence of such NPNs in real utterances; verbal-nexus NNs are thus the only ones that may be considered as mere stylistic variants of the respective NPNs. Attributive NNs, i.e. those where the relationship of the head with the (literally interpreted) modifier is paraphrasable by an 'is a' relation, cannot in principle have NPN variants, and they actually do not have NPN variants in corpus data. As for appositive NNs and for NNs in which the non-head has a bound meaning, the existence of NPN was expected to be very unlikely and indeed, only occasional synonymous NPNs were identified in the data.

The case of subordinate ground NNs proved to be more complex. Although, in theory, most of ground NNs could have an NPN variant, corpus data show that such NPNs tend to be underrepresented in real life utterances. Since many new subordinate ground NNs tend to be organized around paradigmatic families with the same N1 or the same N2 component, we have put forward the hypothesis that such paradigmatic regularity underpins a progressive formation of 'niches' in which new subordinate ground NNs win the competition with NPNs. Moreover, it has been demonstrated that mechanisms of constructionalization even give rise to new subpatterns of NNs for which NPN variants are hardly conceivable. In other terms, these "adaptive changes in the competing species" (Aronoff 2016) may lead to a progressive functional differentiation of the NN and the NPN patterns.

In CM terms, the constructionalization of new subpatterns that underlies the progressive "division of labour" between French GROU-NNs and NPNs consists either in component-based generalizations or in subsequent semantic generalizations. Both processes may be observed on concrete examples of diachronic competition in

²⁹ The competition of NNs and NPNs from Table 6 cannot be visualized on GN, since these sequences pertain to topics and styles that are under-represented in Google books.



Google n-grams. Nevertheless, the phenomenon is still too recent and the existing diachronic corpora are not representative enough, so further verification will be necessary in the future.

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

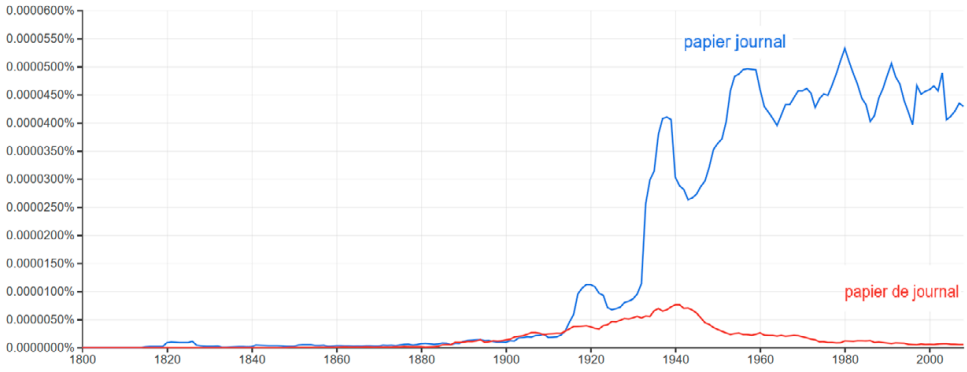
- Allen, M. R. (1978) *Morphological Investigations*. Doctoral Dissertation. University of Connecticut.
- Arnaud, P. J. L. (2003) *Les composés timbre-poste*. Lyon: Presses universitaires de Lyon.
- Arnaud, P. J. L. (2015) Noun-noun compounds in French. In: Müller, P. et al. (eds) *Word formation. An international handbook of the Languages of Europe*, Vol. 3, 673–687. Berlin/Boston: Walter de Gruyter.
- Aronoff, M. (2016) Competition and the lexicon. In: Elia, A., C. Iacobini and M. Voghera (eds) *Livelli di Analisi e fenomeni di interfaccia. Atti del XLVII Congresso Internazionale della Società di Linguistica Italiana*, 39–52. Roma: Bulzoni Editore.
- Baroni, M., E. Guevara and R. Zamparelli (2009) The dual nature of Deverbal Nominal Constructions: Evidence from acceptability ratings and corpus analysis. *Corpus Linguistics and Linguistic Theory* 5/1, 27–60.
- Baroni, M., E. Guevara and V. Pirrelli (2009) Sulla tipologia dei composti N+N in italiano: principi categoriali ed evidenza distribuzionale a confronto. In: Benatti, R., G. Ferrari and M. Mosca (eds) *Linguistica e modelli tecnologici di ricerca (Atti del 40esimo Congresso della Società di Linguistica Italiana)*, 73–95. Roma: Bulzoni.
- Bauer, L. (2001) Compounding. In: Haspelmath, M. (ed) *Language Typology and Language Universals*, 695–707. The Hague: Mouton de Gruyter.
- Benveniste, E. (1974) Formes nouvelles de la composition nominale. *Problèmes de linguistique générale* 2, 163–176. Paris: Gallimard.
- Bonami, O. and J. Strnadová (2019) Paradigm structure and predictability in derivational morphology. *Morphology* 29/2, 167–197.
- Booij, G. E. (2010) *Construction morphology*. Oxford: Oxford University Press.
- Booij, G. E. (2016) Construction morphology. In: Hippisley, A. and G. Stump (eds) *The Cambridge Handbook of Morphology*, 424–448. Cambridge: Cambridge University Press.
- Dardano, M. (2009) *Costruire parole*. Bologna: Il Mulino.
- 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.
- Fradin, B. (2009) IE, Romance: French. In: Lieber, R. and P. Štekauer (eds) *The Oxford handbook of compounding*, 417–435. Oxford: Oxford University Press.
- 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.
- Gaeta, L. and M. Angster (2019) Stripping paradigmatic relations out of the syntax. *Morphology* 29/2, 249–270.
- Gardani, F., F. Rainer and H. C. Luschützky (2019) Competition in Morphology: A Historical Outline. In: Rainer, F., F. Gardani, W. U. Dressler and H. C. Luschützky (eds)

- Competition in Inflection and Word-Formation*, 3–36. Cham: Springer International Publishing.
- Google Books — Available at <https://books.google.com>.
- Google n-grams — Available at <https://books.google.com/ngrams>.
- Haspelmath, M. and A. D. Sims (2010) *Understanding Morphology*. London: Hodder Education.
- 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.
- Lieber, R. and S. Scalise (2006) The Lexical Integrity Hypothesis in a new theoretical universe. *Lingue e linguaggio* V/1, 7–32.
- Lieber, R. (2016) Compounding in the lexical semantic framework. In: Ten Hacken, P. (ed) *The Semantics of Compounding*, 38–53. Cambridge: Cambridge University Press.
- Lindsay, M. and M. Aronoff (2013) Natural selection in self-organizing morphological systems. In: Montermini, F., G. Boyé and J. Tseng (eds) *Morphology in Toulouse: Selected Proceedings of Décembrettes 7*, 133–153. Munich: Lincom Europa.
- Marchand, H. (1969) *The categories and types of present-day English word-formation*. München: Verlag C. H. Beck.
- Moyna, M. I. (2011) *Compound words in Spanish*. Amsterdam: John Benjamins.
- Noailly, M. (1990) *Le substantif épithète*. Paris: PUF.
- Radimský, J. (2015) Noun+Noun Compounds in Italian. A corpus-based study. *České Budějovice: Jihočeská univerzita, edice Epistémé*.
- Radimský, J. (2018a) Does French have verbal-nexus Noun+Noun compounds? *Linguisticae Investigationes* 41/2, 214–224.
- Radimský, J. (2018b) Inflection of binominal ATAP compounds in French and Italian: a paradigmatic account. *Lingue e linguaggio* XVII/2, 261–272.
- Radimský, J. (2019) Les composés N-N de subordination: un paradigme émergent. *Studia Romanica Posnaniensia* 46/1, 167–180.
- Radimský, J. (2020) A paradigmatic account of lexical innovation: the role of repeated components in French N+N compounds. In: Audring, J., N. Koutsoukos and C. Manouilidou (eds) *Rules, patterns, schemas and analogy*. Online proceedings from the 12th Mediterranean Morphology Meeting. Available at <https://mmm.library.upatras.gr/> [last accessed in May 2020].
- Rainer, F. and C. Buridant (2015) From Old French to Modern French. In: Müller, P. O. (ed) *Word-formation: an international handbook of the languages of Europe*. Vol. 3, 1975–2000. Berlin/Boston: Walter de Gruyter.
- Samvellián, P. (1995) *Les nominalisations en français: arguments sémantiques et actants syntaxiques*. Thèse de doctorat, Paris: Université Paris VII.
- Scalise, S. and A. Bisetto (2009) The classification of compounds. In: Lieber, R. and P. Štekauer (eds) *The Oxford handbook of compounding*, 34–53. Oxford: Oxford University Press.
- Strnadová, J. (2008) *N de N — analyse morphologique et syntaxique*. MA thesis, Charles University in Prague.
- Strnadová, J. (2014) *Les réseaux adjectivaux: Sur la grammaire des adjectifs dénominaux en français*. PhD thesis, Charles University in Prague & Université Paris Diderot.
- Tesnière, L. (1959) *Éléments de syntaxe structurale*. Paris: C. Klincksieck.
- Traugott, E. C. and G. Trousdale (2013) *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Van Goethem, K. and D. Amiot (2019) Compounds and multi-word expressions in French. In: Schlücker, B. (ed) *Complex lexical units: compounds and phrases*, 127–152. Berlin: De Gruyter.
- Villoing, F. (2012) French compounds. *Probus* 24/1, 29–60.

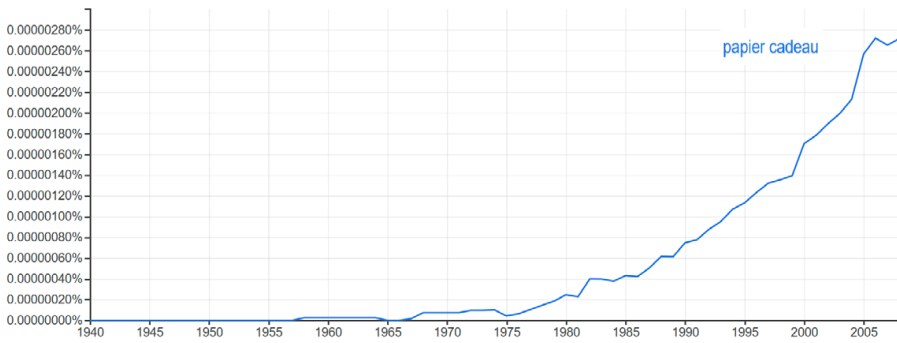




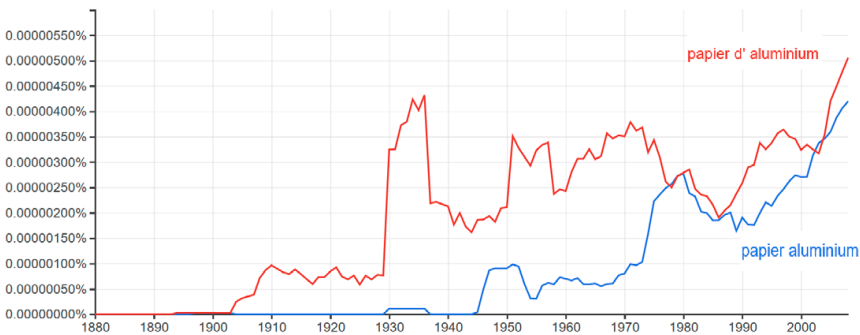
ANNEXES



1.1. Diachronic competition between *papier journal* and *papier de journal*



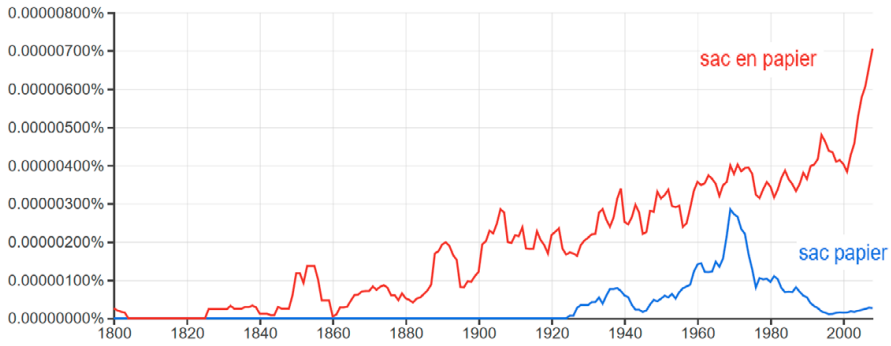
1.2. Diachronic curve for *papier cadeau* (the respective NPN is not attested in Google n-grams)



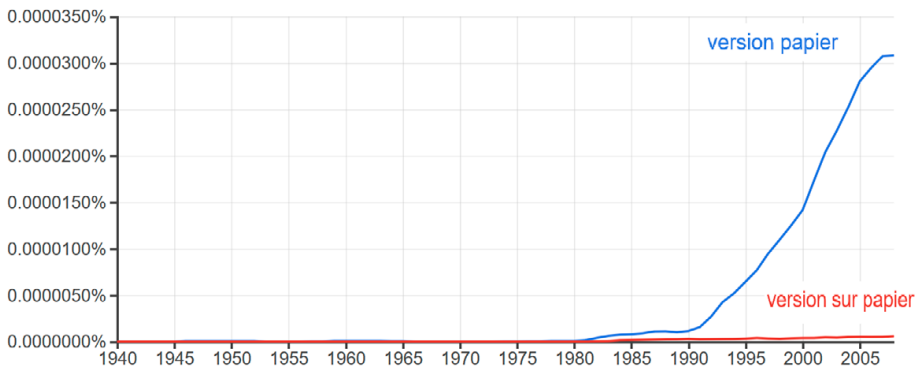
1.3. Diachronic competition between *papier aluminium* and *papier d'aluminium*



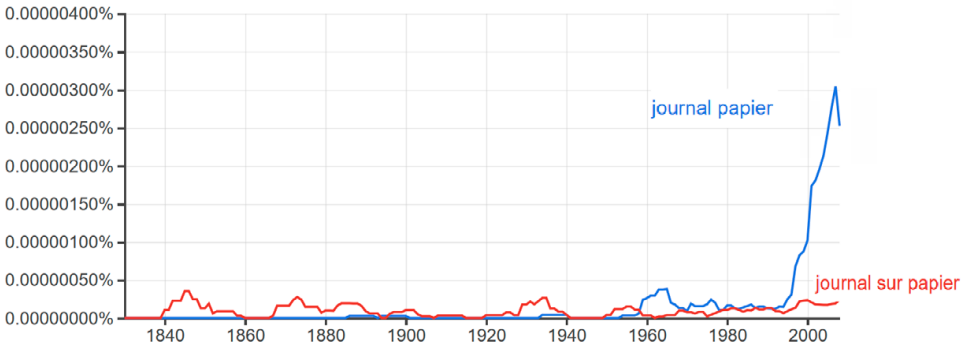
2.1. Diachronic competition between *impression papier* and *impression sur papier*



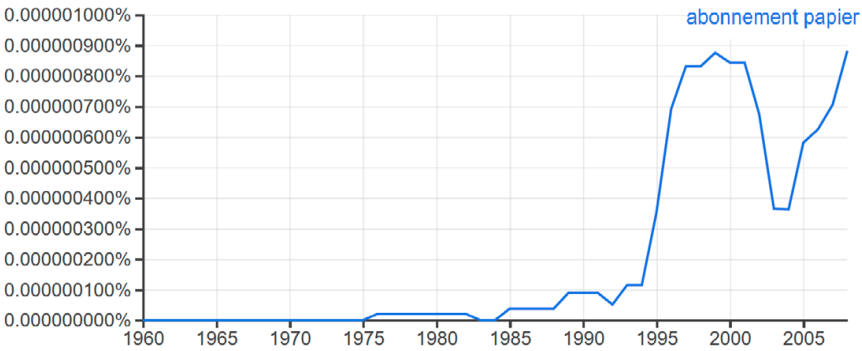
2.2. Diachronic competition between *sac papier* and *sac en papier*



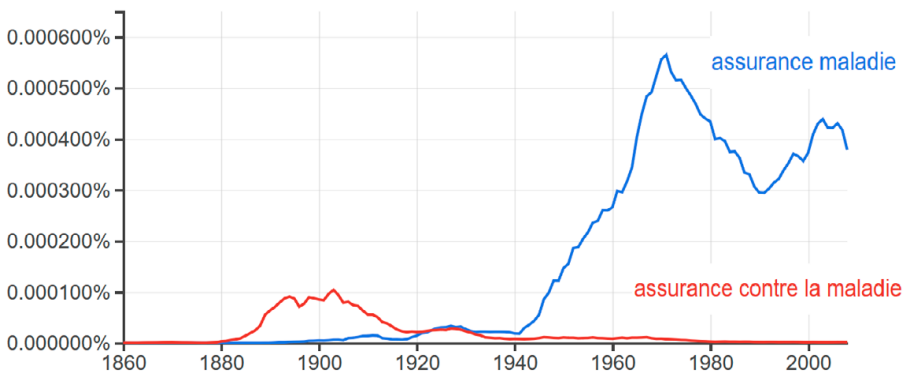
2.3. Diachronic competition between *version papier* and *version sur papier*



2.4. Diachronic competition between *journal papier* and *journal sur papier*



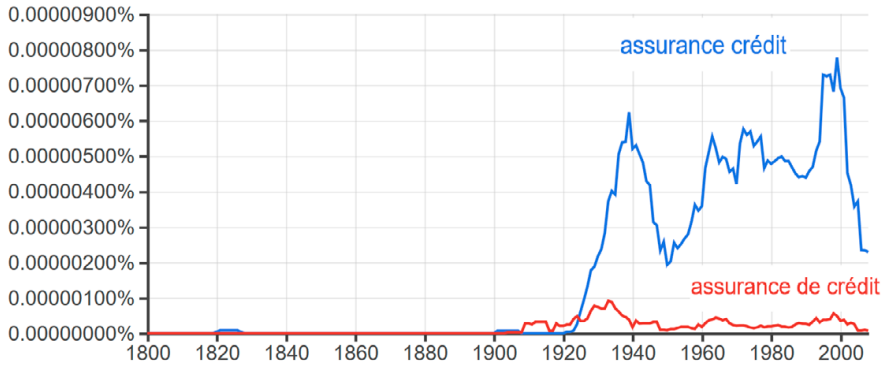
2.5. Diachronic competition between *abonnement papier* and *abonnement sur papier*



3.1. Diachronic competition between *assurance maladie* and *assurance contre la maladie*



3.2. Diachronic competition between *assurance vie* and *assurance sur la vie*



3.3. Diachronic competition between *assurance crédit* and *assurance de crédit*

N1	N2
assurance 'insurance'	<i>accident</i> — 'accident'
	<i>décès</i> — 'death'
	<i>invalidité</i> — 'disability'
	<i>maladie</i> — 'disease' ('health insurance')
	<i>chômage</i> — 'unemployment'
	<i>vie</i> — 'life'
	<i>santé</i> — 'health'
	<i>auto</i> — 'car'
	<i>bagages</i> — 'luggage'
	<i>habitation</i> — 'home'
	<i>crédit</i> — 'credit'

TABLE 7 — NNs capturing the conceptual classification for *assurance* ('insurance')



N1	N2
coin 'section, area of a house'	<i>cuisine</i> — 'kitchen' (<i>'kitchenette'</i>)
	<i>séjour</i> (<i>'living /room/'</i>)
	<i>télé</i> — 'TV'
	<i>cheminée</i> — 'chimney'
	<i>lavabo</i> — 'lavatory'
	<i>repas</i> — 'meal'
	<i>café</i> — 'coffee'
	<i>repos</i> — 'rest'
<i>enfant</i> — 'child'	
<i>fumeur</i> — 'smoker'	

TABLE 8 — NNs capturing the conceptual classification for *coin* ('section, area of a house or store')

N1	N2
pause 'break'	<i>café</i> — 'coffee'
	<i>cigarette</i> — 'cigarette'
	<i>repas</i> — 'meal'
	<i>sandwich</i> — 'sandwich'
	<i>thé</i> — 'tea'
	<i>goûter</i> — 'snack'
	<i>lecture</i> — 'reading'
<i>détente</i> — 'relaxation'	
<i>midi</i> — 'midday'	

TABLE 9 — NNs capturing the conceptual classification for *pause* ('break')

N1	N2
tarif 'rate, tariff'	<i>enfant</i> — 'child'
	<i>abonné</i> — 'subscriber'
	<i>auto</i> — 'car'
	<i>journée</i> — 'day'
	<i>discount</i> — 'discount'
<i>été</i> — 'summer'	

TABLE 10 — NNs capturing the conceptual classification for *tarif* ('rate, tariff')

Jan Radimský

Department of Romance Studies

Faculty of Arts, University of South Bohemia in České Budějovice

Branišovská 31a, 37005 České Budějovice

ORCID ID: 0000-0002-9016-9293

radimsky@ff.jcu.cz