

Les composés *N-N* de subordination : un paradigme émergent¹

French subordinate *N-N* compounds: an emerging paradigm

Jan Radimský

Université de Bohême du Sud, České Budějovice

radimsky.jan@gmail.com

Abstract

This paper aims at examining the causes of the emergence of French subordinate Noun-Noun compounds. It is well known that the Noun-Noun pattern in French remains marginal compared to other lexicogenic processes, especially N-PREP-N or N-A, and it is supposed that its appearance as well as its progressive development took place during the last two centuries (19th-20th). The aim of this paper is to examine more in detail *when* and *why* French Noun-Noun structures emerge. As for the first question, empirical data from the Frantext corpus allow to hypothesize that both the type and the token frequency of French Noun-Noun compounds remain stable since the thirties of the nineteenth century until the end of the Second World War and that after this period, especially during the sixties, it begins to grow exponentially. Contrary to Arnaud's estimate (2003, p. 141), no significant change in frequency or productivity was observed around the middle of the 19th century. As for the second question, the author claims that the emergence of subordinate N-N compounds was triggered by an increase in the productivity of the attributive N-N compounds, for which there is no competitive pattern in French. The theoretical rationale of this hypothesis is anchored in paradigmatic approaches to word formation, with specific reference to the formalization made according to the Construction Morphology framework (Booij, 2010).

Keywords: Noun-Noun compounds, French, Construction Morphology, French compounds

¹ La présente contribution a été réalisée grâce au soutien financier de l'Agence nationale tchèque pour la recherche (GAČR), attribuée dans le cadre du projet GA17-17253S – « Les composés N-N en français contemporain ».

1. INTRODUCTION

La présente contribution a pour objet d'examiner les motifs de l'émergence des composés français de subordination relevant du patron N-N tels que *exposition photos*, *abonnement internet*, *timbre-poste* ou *sauce-tomate*. Il est bien connu qu'en français – comme dans les langues romanes en général – le patron N-N reste marginal par rapport à d'autres procédés lexicogénétiques, notamment les « synapses » N-PREP-N (*pomme de terre*) ou N-A (*voie lactée*), et il est supposé que son apparition ainsi que son développement progressif a eu lieu notamment au cours des deux derniers siècles (19^e-20^e). Néanmoins, une étude à la fois empirique et théorique reste nécessaire afin de dévoiler en détail les étapes diachroniques ainsi que les motifs de ce processus ; autrement dit, il faudrait être en mesure de dire plus exactement *quand* et *pourquoi* les composés N-N de subordination français apparaissent. Ces deux questions feront l'objet de la présente contribution.

En ce qui concerne le volet empirique, les données diachroniques qui rendront compte des étapes de l'apparition des composés N-N à partir du début du 19^e siècle seront extraites du corpus Frantext. Quant au volet théorique, nous proposerons l'hypothèse selon laquelle l'émergence des composés N-N de subordination a été déclenchée par une hausse de la productivité des composés N-N attributifs, pour lesquels il n'y a pas de patron concurrentiel en français. Confirmée empiriquement par les données de Frantext, l'hypothèse en question trouve son ancrage théorique dans les approches paradigmatiques à la formation des mots, dont notamment la Morphologie constructionnelle (Booij, 2010).

L'article sera organisé comme suit : la section 2 introduira les types majeurs de composés N-N français ainsi que leurs origines possibles selon les études antérieures (notamment Arnaud, 2003 ; Rainer-Buridant, 2015). Dans la section 3, les relations entre les différentes constructions N-N seront analysées en termes de Morphologie constructionnelle, ce qui nous permettra de formuler l'hypothèse-clé sur l'émergence des composés N-N de subordination. La section 4 sera consacrée à l'analyse des données extraites du corpus Frantext.

2. SOURCES POSSIBLES DES COMPOSES N-N

En ce qui concerne la typologie des composés N-N français, nous allons nous référer à la classification générale des composés proposée par Bisetto et Scalise (Bisetto-Scalise, 2005, modifiée et mise à jour dans Scalise-Bisetto, 2009), dont l'application aux composés N-N romans a été discutée en détail par Radimský (2015, p. 85-160)². Aux fins de la présente analyse, trois grands types de structures

² Pour une application de la typologie de Bisetto et Scalise aux composés N-N français voir p. ex. Villoing (2012).

N-N vont être distingués, à savoir les composés attributifs (1), les composés de coordination (2) et les composés de subordination (3).

(1) *bourgeois gentilhomme*

(2) *boulangerie-pâtisserie*

(3) *timbre-poste*

En principe, les composés attributifs (1) sont ceux dont les composants entretiennent une relation du type tête-attribut, ce qui fait qu'ils sont susceptibles d'une paraphrase par une structure copulative du type (4).

(4) *ce bourgeois gentilhomme* → *ce bourgeois EST UN gentilhomme*

Les composés de coordination (2) sont une variante symétrique des composés attributifs, dans le sens où la paraphrase par une structure copulative est réversible (5a-b). Pour cela, il est nécessaire que les composants désignent des concepts du même type situés au même niveau de la hiérarchie conceptuelle. En effet, si *boulangerie* désigne une *activité commerciale ayant trait à la fabrication et à la vente du pain*, le mot *pâtisserie* fait référence à une *activité commerciale ayant trait à la fabrication et à la vente des gâteaux*³.

(5a) *cette boulangerie-pâtisserie* → *cette boulangerie EST UNE pâtisserie*

(5b) *cette boulangerie-pâtisserie* → *cette pâtisserie EST UNE boulangerie*

Les composés de subordination (3) sont ceux dotés d'une tête et d'un modifieur au sens large du terme (i.e. d'un nom subordonné à la tête). Si les relations sémantiques entre la tête et son modifieur peuvent être très variées (cf. Arnaud, 2003), l'étude de leur variabilité n'est pas pertinente pour la présente analyse. Il suffira donc de définir les composés N-N de subordination de manière négative en tant que composés dotés d'une tête à la position N1 qui ne relèvent ni des composés attributifs (1), ni des composés de coordination (2).

Cette typologie simplifiée laisse volontairement de côté certains types marginaux, tels que les composés de coordination exocentriques (dvandva) comme *histoire-géographie* (Villoing, 2012, p. 45) ou les composés dont la tête se situe à droite (i.e. en position N2), comme dans *[[auto]_{N1}[route]_{N2}]_N*.

Quant à l'origine des différents types de composés N-N, Rainer et Buridant (2015 : 1978) considèrent que les composés de coordination (*boulangerie-pâtisserie*) représentent une extension du modèle des structures attributives (*bourgeois gentilhomme*) qui ont toujours été présentes en français ; notons que cette hypothèse est parfaitement compatible avec la typologie des composés présentée ci-dessus, où

³ Nous ne faisons pas recours à la paraphrase par la conjonction « et » qui est réservée aux composés de coordination exocentriques du type « dvandva » comme *histoire-géographie*.

les composés de coordination sont considérés comme une simple variante des composés attributifs. Par contre, le patron des composés de subordination (*timbre-poste*) aurait selon Rainer et Buridant (2015 : 1978) des sources « variées » et son essor ne devrait pas être attribué uniquement à l'influence anglo-saxonne. En effet, au moins quatre hypothèses différentes sur l'origine possible des composés N-N de subordination français ont été discutées dans la littérature, dont (a) le génitif de juxtaposition, (b) l'instabilité de la préposition dans les synapsies, (c) l'influence des langues germaniques et (d) l'analogie avec les structures attributives. Nous allons présenter brièvement les différentes hypothèses en nous référant notamment à la synthèse faite par Arnaud (2003, p. 136-141).

Le génitif de juxtaposition (a) qui consiste à ajouter au nom-tête un complément nominal au cas régime a engendré les structures telles que (6a-c). Remarqué déjà par Darmesteter (1874, p. 48-49), le génitif de juxtaposition correspondait à une construction syntaxique régulière, ce qui fait que sa ressemblance au patron N-N contemporain n'est qu'apparente (Rainer-Buridant, 2015 : 1977). De plus, cette construction n'est plus disponible à partir du 15^e siècle et elle semble limitée au cas où le N2 (au cas régime) est un nom propre ou un nom de personne avec une référence définie⁴. De ce fait, Arnaud (2003, p. 137) considère que son rôle dans le développement du patron N-N de subordination en français moderne reste très marginal.

(6a) *ostel Dieu* (« Hôtel Dieu »)

(6b) *bainmarie*

(6c) *li filz le rei* (« le fils du roi »)

Une hypothèse originale (b) formulée par Hatcher (1946) est basée sur l'observation du fait que le champ d'application de la préposition *à* se réduit au profit d'autres prépositions depuis l'ancien français. Par conséquent, on assisterait à une instabilité de la préposition dans les synapsies N-PREP-N (7a-b), ce qui entraînerait l'omission complète de la préposition entre les deux noms.

(7a) *wagon de/à marchandises*

(7b) *boîte d'/à outils*

Nous rejoignons l'opinion d'Arnaud (2003, p. 137) qui considère cette hypothèse, illustrée seulement sur une vingtaine de composés, peu convaincante du point de vue théorique aussi bien que du point de vue empirique⁵.

⁴ En utilisant la terminologie propre à la grammaire générative, on pourrait affirmer que le N2 dans les constructions du type *li filz le rei* est un syntagme déterminatif (DP, *determiner phrase*), tandis que le N2 dans les composés N1-N2 est un syntagme nominal (NP, *noun phrase*).

⁵ Cf. également à ce propos A. Darmesteter (1877, p. 159-160).

L'hypothèse selon laquelle l'influence directe de l'anglais ou d'une autre langue germanique (c) a joué un rôle majeur dans la création et l'expansion du patron N-N de subordination français partage les linguistes : ses partisans (Darmesteter (1874), Rohlf, Guiraud, Grevisse, Hatcher) sont aussi nombreux que ses opposants (Darmesteter (1877), Guilbert, Picone, Rainer-Buridant, Arnaud). Cependant, il est symptomatique dans ce panorama que les auteurs les plus récents (Picone, 1996 ; Arnaud, 2003 ; Rainer-Buridant, 2015) relèvent du second groupe. En effet, si l'influence de l'anglo-américain, devenue incontestable à partir de la seconde moitié du 20^e siècle, a pu jouer un certain rôle dans l'expansion du patron N-N, les données empiriques montrent que l'émergence des composés N-N de subordination avant cette période ne peut pas être attribuée à l'importation d'un modèle morphologique étranger, du moins pas uniquement. Arnaud (2003, p. 138) constate à ce propos que les composés N-N de subordination « antérieurs à 1850 sont massivement indigènes, et peu nombreux sont ceux qui sont des calques évidents non pas même du seul anglais, mais d'une langue germanique ».

La dernière hypothèse (d), selon laquelle le patron des composés N-N de subordination se serait formé par analogie à partir du patron des structures attributives, ne semble pas avoir une répercussion importante dans la littérature. L'idée a été formulée en passant par A. Darmesteter (1874, p. 138-139) dans une note en bas de page, où il est affirmé que la composition par apposition « semble devenir désormais comme un moule dans lequel l'analogie tend à faire entrer nombre de composés dont l'un des termes est le régime de l'autre ». Si Darmesteter lui-même n'a pas essayé de donner une description plus précise du mécanisme de cette analogie, Hatcher (1946, p. 216-217) en propose une interprétation directe qui prévoit une extension du rapport logique entre les constituants du composé. Selon elle, l'extension analogique prévoirait par exemple le passage d'une interprétation attributive (8a) vers une interprétation locative (8b) qui, à son tour, devrait aboutir à une formation libre des composés N-N à rapport locatif tels que (9).

(8a) *roman feuilleton* – *roman qui est un feuilleton*

(8b) *roman feuilleton* – *roman qui se trouve dans le feuilleton*

(9) **maison campagne*, **lit-hôpital*, **feux-rampe*, **fille-trottoir*

L'agrammaticalité des composés en (9) infirme, selon Hatcher, l'hypothèse analogique de Darmesteter.

Nous sommes convaincu que parmi les quatre hypothèses (a-d) sur l'origine du patron N-N de subordination, celle de Darmesteter (d) est la plus plausible, mais le mécanisme du développement du patron de subordination à partir du patron attributif doit être formalisé différemment. Dans la section suivante, une formalisation basée sur la Morphologie constructionnelle sera proposée.

3. RELATIONS ENTRE LES CONSTRUCTIONS N-N

En termes de Morphologie constructionnelle (Booij, 2010, 2016), les patrons compositionnels peuvent être décrits en tant que *constructions*, i.e. comme des couples forme-sens formalisables par un *schéma* constructionnel. Si, dans une perspective ascendante, les différentes constructions représentent des abstractions faites à partir des composés qui existent dans le lexique, de nouveaux composés peuvent être formés, dans une perspective descendante, en tant qu'instances des différentes constructions. Ainsi, on peut affirmer que les trois grands types de composés français N-N présentés dans la section 2 correspondent chacun à une construction, formalisable par les schémas respectifs en (10a-c).

(10a) Composé N-N attributif (*bourgeois gentilhomme*)

$[N_i N_j]_{NK} \leftrightarrow [SEM_i \text{ est un(e) } SEM_j]_k$

(10b) Composé N-N de coordination (*libraire-éditeur*)

$[N_i N_j]_{NK} \leftrightarrow [SEM_i \text{ est un(e) } SEM_j]_k \leftrightarrow [SEM_j \text{ est un(e) } SEM_i]_k$

(10c) Composé N-N de subordination (*timbre-poste*)

$[N_i N_j]_{NK} \leftrightarrow [SEM_i \text{ REL } SEM_j]_k$

Les composés N-N de coordination (10b) représentent une variante symétrique des composés attributifs, ce qui fait qu'ils sont potentiellement réversibles (*éditeur-libraire*) avec un glissement de sens plus ou moins prononcé⁶. Dans le schéma des composés de subordination (10c), l'abréviation REL fait référence à une vaste gamme de relations sémantiques⁷.

Il est important de noter qu'en morphologie constructionnelle, les constructions sont organisées hiérarchiquement (Booij, 2016, p. 430-433), ce qui fait que les constructions qui partagent des propriétés communes forment une construction plus abstraite (en perspective ascendante) et en même temps, les constructions de rang inférieur héritent les propriétés de la construction de rang supérieur (en perspective descendante). Ainsi, les constructions formalisées par les schémas (10a-c) représentent des instances d'une construction de rang supérieur formalisable par le schéma (11).

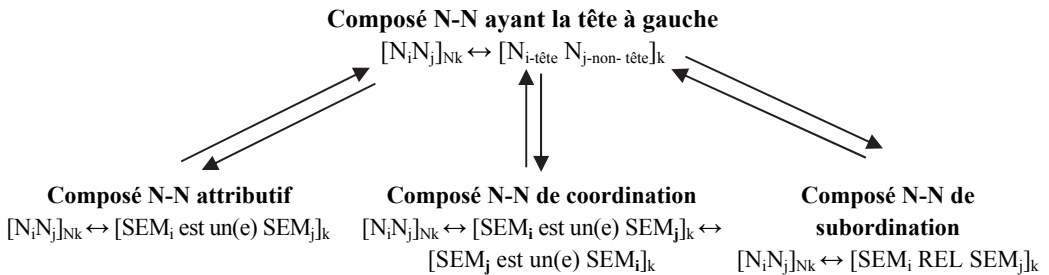
(11) Composé N-N ayant la tête à gauche

$[N_i N_j]_{NK} \leftrightarrow [N_{i-tête} N_{j-non-tête}]_k$

⁶ Voir une discussion détaillée dans Radimský (2015, p. 102-112).

⁷ Nous pourrions ajouter à ce panorama les composés N-N appositifs (*oiseau-mouche*) formalisables par la construction $[N_i N_j]_{NK} \leftrightarrow [SEM_i \text{ est } \mathbf{comme} \text{ un(e) } SEM_j]_k$. Situés à cheval entre les composés attributifs et les composés de subordination, les composés appositifs sont parfois assimilés aux composés attributifs (Bisetto-Scalise, 2005) ou aux composés de subordination (Arnaud, 2003) ou, dans d'autres cas, ils sont présentés comme un sous-type spécifique du groupe attributif-appositif (ATAP) (Scalise-Bisetto, 2009 ; Radimský, 2015).

Les constructions (10a-c) qui correspondent aux trois différents types de composés ne sont donc pas indépendantes : elles sont dans un rapport hiérarchique avec la construction (11), un rapport qui peut être schématisé comme suit.



Parmi les constructions de rang inférieur, celle qui correspond aux composés N-N attributifs (*bourgeois gentilhomme*) est particulière en deux aspects ; du point de vue diachronique, elle a toujours été disponible en français et, en perspective onomasiologique, elle n'a pas de patron concurrentiel. En effet, à part quelques cas exceptionnels comme les noms des mois (*mois de mai*), les toponymes (*ville de Paris*) ou les insultes (*crétin de Paul*) réalisés par des constructions dont la forme correspond à la structure N-PREP-N, le rapport attributif revêt toujours la forme N-N en français⁸. Une présence accrue des composés N-N attributifs – traduite par une hausse de la productivité des types dans le lexique et/ou par une présence croissante des tokens respectifs dans les textes – entraîne donc non seulement l'actualisation de la construction des composés N-N attributifs (10a), mais également l'actualisation de la construction plus générale qui correspond aux composés N-N ayant la tête à gauche (11). Les nouveaux composés N-N de coordination et de subordination sont donc des instances de la construction (11) où la contrainte particulière qui limite le rapport entre les membres du composé au rapport attributif (10a) n'est plus respectée. Ceci dit, les nouveaux composés N-N de subordination peuvent entretenir en principe n'importe quel type de lien sémantique entre les composants. L'expansion du modèle reste cependant fortement freinée par la présence de nombreux patrons (i.e. constructions) à forme N-PREP-N qui sont en concurrence directe avec les N-N de subordination. Ainsi, les N-N de subordination se propagent très lentement, ce qui se traduit par l'agrammaticalité des constructions comme (9) signalées par Hatcher (1946). Par contre pour les composés de coordi-

⁸ En tchèque par exemple, le titre de la pièce de Molière est traduit par une structure $N_{NOM}-N_{INS}$ (*měšťák šlechticem*) marquée casuellement, c'est-à-dire que le nom-tête est au nominatif et le nom-attribut en instrumental.

nation (10b), le même type d'inhibition par des constructions concurrentielles ne se manifeste pas, ce qui se traduit par le fait que le schéma (10b) peut être instancié librement, selon le *open choice principle* de Sinclair.

4. ANALYSE DES DONNEES DU FRANTEXT

4.1. Rassemblement des données

La plus grande collection de composés N-N de subordination, rassemblée par Arnaud (2003), compte 921 types, dont 112 sont attestés avant l'année 1850. Cette base de données a été construite à partir de dictionnaires de langue ainsi que de textes authentiques (littéraires ou non), mais l'année d'apparition des différents types n'a pas pu être enregistrée. Par conséquent, Arnaud (2003, p. 141) affirme que « Il n'est pas possible avec les datations dont on dispose d'établir une courbe de l'apparition des nouveaux CTP [= 'composés timbre-poste', i.e. composés N-N de subordination], mais il semble que leur multiplication qui frappait Darmesteter en 1891 et Lombard en 1930 s'accélérait en fait depuis le milieu du XIXe siècle ». L'auteur rappelle cependant que toute conclusion doit être tirée avec prudence, car pour plusieurs raisons, les données en question pourraient représenter seulement « la partie émergée de l'iceberg » (Arnaud, 2003, p. 139)⁹.

Afin de rendre compte de la dimension diachronique de l'apparition et de l'usage des composés N-N, nous avons fait appel au Frantext, un corpus diachronique représentatif des textes littéraires. La partie balisée du Frantext¹⁰, qui a servi de base pour l'extraction des données, couvre les textes provenant de la période entre 1830 jusqu'à 1999 et atteint une taille totale de 1,27x10⁸ positions (soit 127 millions de mots). C'est à partir de ces textes que nous avons extrait la base des données « Binominaux du Frantext » ; cette dernière contient les concordances complètes des séquences N-N apparues dans l'une des deux formes graphiques typiques pour les composés N-N, c'est-à-dire celle où les composants sont séparés par un trait d'union (p. ex. *commis-voyageur*) et celle où ils sont séparés par un espace (p. ex. *commis voyageur*). Dans un premier temps, l'extraction des formes du premier type a été basée uniquement sur la présence du trait d'union (en d'autres

⁹ Selon l'auteur, plusieurs indices laissent croire que les sources écrites ne reflètent pas l'étendue réelle des données. On peut supposer que les composés N-N de subordination relèvent d'un registre moins formel qui est censuré à l'écrit (en effet, certains composés N-N attestés dans les textes sont enregistrés sous forme N-PREP-N dans les dictionnaires d'époque) et que de nombreux composés N-N relèvent également du registre de la langue de spécialité (c'est pourquoi de nombreux exemplaires proviennent du *Dictionnaire universel* de Furetière de 1690 qui, à la différence du Dictionnaire de l'Académie, enregistre de nombreux termes scientifiques et techniques).

¹⁰ La présente recherche fait référence à la version de Frantext qui a été disponible jusqu'au printemps 2018.

termes, toutes les séquences de lettres qui contenaient un trait d'union ont été extraites), tandis que l'extraction des formes du second type se basait sur le balisage des parties du discours (i.e. uniquement les séquences « NOM + ESPACE + NOM » ont été extraites). La base des données « Binominaux du Frantext » comprend, pour chaque concordance, un contexte de 5 positions des deux côtés, la source et l'année de sa publication. Au total, elle comprend 195 330 binominaux potentiels (tokens), dont deux tiers relèvent de la version graphique avec le trait d'union.

Comme la version brute de la base de données contenait un taux de bruit considérable, un filtrage automatique ainsi que manuel s'imposait. Dans un premier temps, les données ont été confrontées avec la liste des noms communs contenue dans la base des données GLAFF (Hathout, Sajous, Calderone, 2014). Cette étape « automatique » a permis d'identifier les formes en position N1 et N2 qui peuvent effectivement correspondre aux formes d'un nom commun français et d'attribuer la forme de base (lemma), le genre et le nombre pour chaque mot candidat. Ensuite, les échantillons correspondant aux différents types de composés N-N ont dû être identifiés à la main.

L'identification manuelle des composés N-N attributifs est relativement simple dans la mesure où même un contexte restreint de +/- 5 mots, enregistré dans la base « Binominaux du Frantext », permet en général de détecter ce type de composé. Comme de nombreux composés attributifs désignent des êtres humains, une attention particulière a été accordée aux composants qui se terminent en « -eur » ou « -iste ». Au total, 1070 composés attributifs (types) qui correspondent à 3791 tokens ont été identifiés manuellement. La même procédure a été utilisée pour les composés de coordination, dont un échantillon de 204 types (soit 493 tokens) a été rassemblé.

Par contre, les composés N-N de subordination sont particulièrement difficiles à identifier à partir d'un contexte restreint. De ce fait, nous avons préféré nous référer à la liste des 921 composés de subordination identifiés par Arnaud (2003) ; parmi eux, 299 types (soit 2384 tokens) apparaissent effectivement dans la base de données « Binominaux du Frantext ».

Un aperçu quantitatif des données rassemblées est donné dans le tableau 1. Comme les composés de coordination sont parfois difficiles à discerner des composés attributifs, les deux types ont été fusionnés en un seul groupe. Au total, l'échantillon comprend 1573 types (soit 6668 tokens).

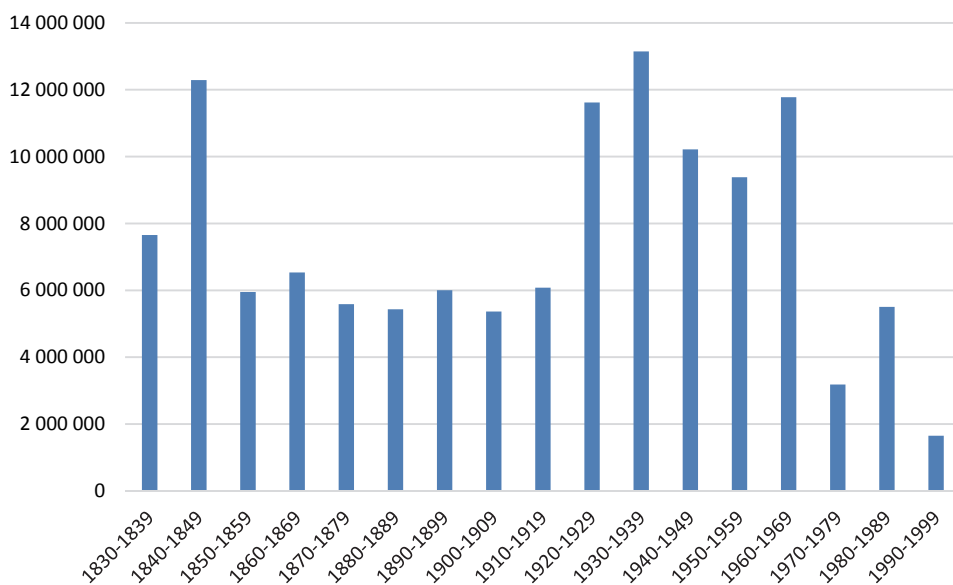
Tableau 1. Composés N-N dépouillés du corpus Frantext

	Composés N-N attributifs & de coordination	Composés N-N de subordination	Total
Types	1274	299 (19%)	1573
Tokens	4284	2,384 (36%)	6668
Rapport token/type	3,4	8,0	4,2

Le fait que les composés de subordination présentent un rapport token/type plus élevé (8,0) par rapport au groupe attributif et de coordination (3,4) est probablement à attribuer à la méthode de dépouillement : basé sur l'échantillon de Arnaud (2003), le groupe subordinatif comprend de nombreux composés lexicalisés et enregistrés dans les dictionnaires.

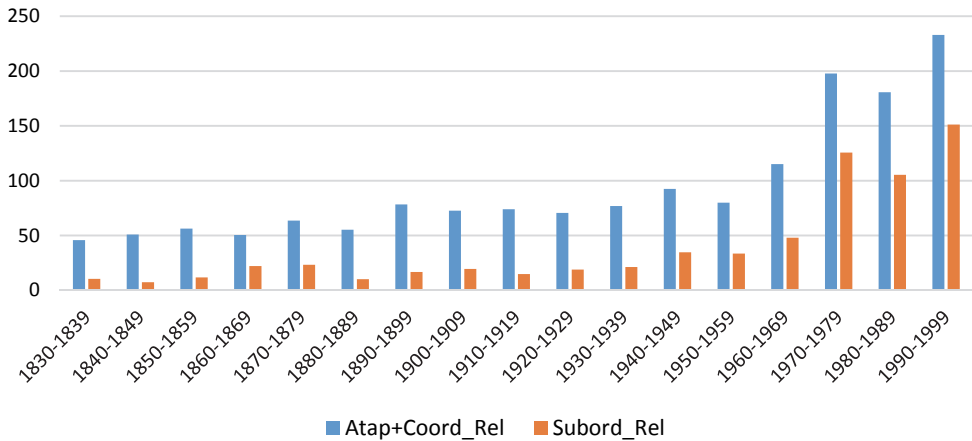
4.2. Présentation des résultats

Pour rendre compte de la dimension diachronique de l'apparition ainsi que de l'usage des composés N-N, les données ont été divisées par tranches de 10 ans. Néanmoins, la taille du corpus Frantext n'est pas la même pour chaque période, comme le montre le graphique 1. Par conséquent, la comparaison des fréquences des composés pour chaque période est basée sur leur fréquence relative, recalculée sur un sous-corpus de 5 millions de mots par tranche.

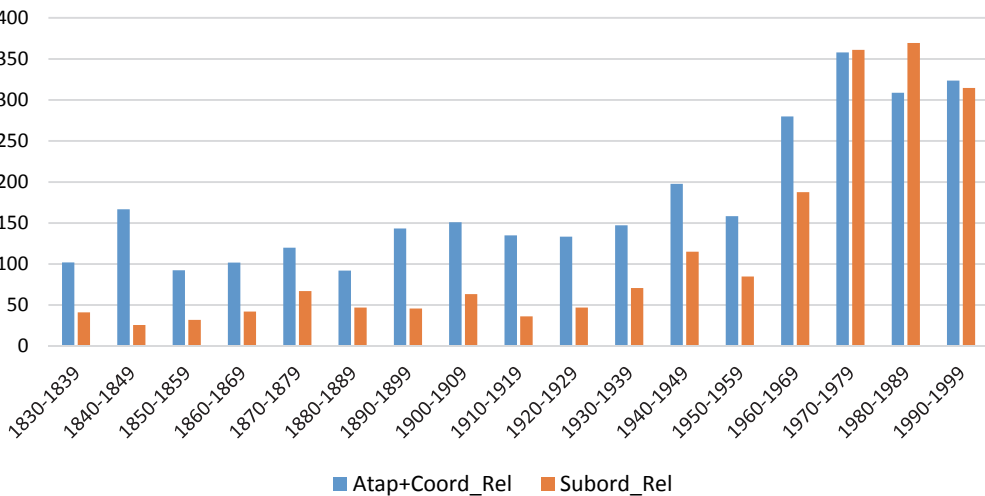


Graph. 1. Taille du corpus Frantext (en nombre de mots) par périodes de 10 ans

Les résultats sont présentés dans les trois graphiques qui suivent. Les graphiques 2 et 3 présentent respectivement la fréquence relative des types et celle des tokens pour les tranches diachroniques de dix ans, allant de 1830 jusqu'à 1999.



Graph. 2. Composés N-N dans Frantext : fréquence relative des types (par 5 millions de mots)

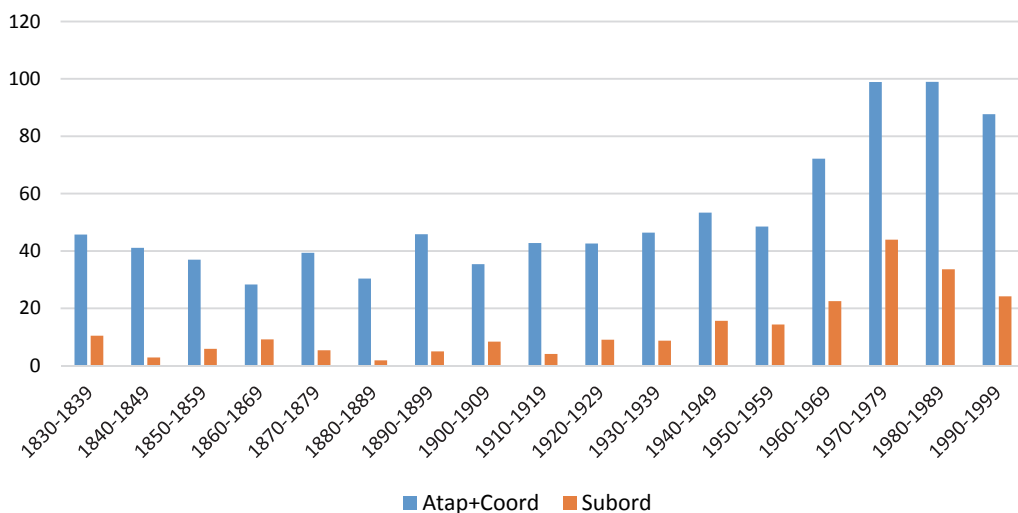


Graph. 3. Composés N-N dans Frantext : fréquence relative des tokens (par 5 millions de mots)

Ces données montrent que la fréquence relative des composés N-N reste plutôt stable dans le corpus pendant plus d'un siècle. Une hausse continue – mais lente – entre les années 1830 et 1949 peut s'observer notamment pour la fréquence des types : le graphique 2 indique en effet qu'au cours de ces 120 années, sa valeur a presque doublé. Par contre, une hausse sensible a lieu dans l'après-guerre, notamment dans les années '60 du XX^e siècle. L'estimation d'Arnaud (2003, p. 141)

qu'une hausse progressive pouvait s'observer déjà depuis la moitié du XIX^e siècle ne semble donc pas se vérifier, le vrai essor intervient un siècle plus tard. En outre, on peut observer qu'à quelques exceptions près, la courbe de fréquence des N-N subordinés suit celle des N-N attributifs et coordinatifs. Ainsi, les données confirment l'hypothèse sur les interrelations entre les différentes constructions N-N présentées dans la section 3.

Si les deux graphiques précédents rendent compte de l'usage des composés N-N dans les textes, la productivité des différents modèles est mieux démontrée par le graphique 4 qui enregistre la fréquence relative des types en se limitant seulement à la date de la première attestation de chaque composé dans Frantext.



Graph. 4. Composés N-N : fréquence relative des types (par 5 millions de mots) selon la 1^{re} attestation dans Frantext

Dans l'ensemble, la courbe de fréquence dans le graphique 4 retranscrit celle des graphiques précédents. Autrement dit, la fréquence de l'usage des composés N-N semble être proportionnelle à la fréquence de la création des nouveaux composés N-N. Seulement pour les deux dernières décennies la fréquence des nouveaux types (graph. 4) décroît, notamment pour les composés de subordination. Cependant, dans les dernières tranches, les données peuvent être moins fiables – d'une part, la taille du corpus est sensiblement plus petite (cf. le graphique 1) et d'autre part, il n'est pas sûr dans quelle mesure l'échantillon d'Arnaud (2003), qui est à la base des composés de subordination dans les graphiques, prend en compte les néologismes des deux dernières décennies du XX^e siècle.

5. CONCLUSION

Malgré certaines limitations empiriques déterminées notamment par une contrainte de genre (textes littéraires) et des contraintes liées au dépouillement du corpus (surtout pour les composés N-N de subordination), la présente étude, basée sur un corpus diachronique représentatif, permet d'avancer de nouvelles hypothèses sur les deux questions posées dans l'introduction, i.e. *quand* et *pourquoi* les composés N-N de subordination commencent à se former et à se développer en français.

Pour répondre à la première question, on peut affirmer que dans les textes littéraires, la fréquence des types ainsi que celle des tokens des composés N-N reste stable depuis les années '30 du XIX^e siècle jusqu'à la fin de la Seconde Guerre mondiale et qu'après cette période, notamment au cours des années '60, elle croît de manière exponentielle. La même allure a été observée pour la courbe de fréquence des nouveaux composés N-N attestés dans le corpus. Contrairement à l'estimation d'Arnaud (2003, p. 141), aucun changement significatif de fréquence ou de productivité n'a été observé vers la moitié du XIX^e siècle.

En ce qui concerne la seconde question, les données montrent que les courbes de fréquence des composés de subordination présentent la même allure que celle des composés attributifs et de coordination – qu'il s'agisse de la fréquence des tokens, des types ou des nouveaux types attestés. Ceci est en conformité avec l'hypothèse selon laquelle les composés N-N de subordination, de coordination et attributifs représentent des instances spécifiques de la même construction générale $[N_i N_j]_{Nk} \leftrightarrow [N_{i-tête} N_{j-non-tête}]_k$ ce qui veut dire, en d'autres termes, que le développement des composés N-N de subordination peut représenter une extension du modèle des composés attributifs. Le fait que tous les types de composés N-N ont connu une hausse exponentielle de productivité depuis les années '60 du XX^e siècle permet d'avancer l'hypothèse que l'influence de l'anglo-américain a pu jouer le rôle de catalyseur dans ce processus.

Afin de vérifier et de préciser ces hypothèses ultérieurement, il serait souhaitable de disposer de corpus représentatifs diachroniques qui couvriraient également d'autres genres (tels les textes de spécialité, les textes administratifs, journalistiques, etc.). En outre, il faudrait être en mesure d'assurer que le balisage du corpus soit fiable. En effet, une simple requête effectuée sur le Google n-gram viewer (basé sur le corpus des Google books français)¹¹ laisse croire que les séquences (i.e. « bigrams ») N-N présentent une courbe de fréquence plus stable tout au long des deux derniers siècles. Néanmoins, si le balisage a été effectué à partir d'un baliseur HMM entraîné sur une seule tranche synchronique des données, ce résultat – qui contraste sensiblement avec nos conclusions basées en grande partie sur les données non balisées – n'est pas surprenant. D'autre part, il serait également souhaitable étendre cette recherche à d'autres langues romanes où les composés N-N apparaissent.

¹¹ Cf. la séquence « _NOUN_ _NOUN_ » recherchée sur : <https://books.google.com/ngrams>.

BIBLIOGRAPHIE

- Arnaud, P. (2003). *Les composés « timbre-poste »*. Lyon : Presses Universitaires de Lyon.
- Bisetto, A., Scalise, S. (2005). « The classification of compounds ». *Lingue e Linguaggio*, 4 (2), 319-332.
- Booij Geert, E. (2010). *Construction Morphology*. Oxford : Oxford University Press.
- Booij Geert, E. (2016). « Construction morphology ». In : A. Hippisley, G. Stump, *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.
- Darmesteter, A. (1877). *De la création actuelle de mots nouveaux dans la langue française et des lois qui la régissent*. Paris : F. Vieweg.
- Frantext, *Corpus de langue française*, <https://www.frantext.fr/> (consulté en mars-avril 2018).
- Hatcher, A.G. (1946). « Le Type Timbre-Poste ». *Word*, 2-3, 216-228.
- Hathout, N., Sajous, F., Calderone, B. (2014). « GLÀFF, a Large Versatile French Lexicon ». Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, 1007-1012.
- Picone, M.D. (1996). *Anglicisms, Neologisms and Dynamic French*. Amsterdam : John Benjamins.
- Radimský, J. (2015). *Noun + Noun Compounds in Italian. A corpus-based study*, edice Epistémé. České Budějovice : Jihočeská univerzita.
- Rainer, F., Buridant, C. (2015). « From Old French to Modern French ». In : P.O. Muller (ed.), *Word-formation: an international handbook of the languages of Europe*, vol. 3. Berlin/Boston : Walter de Gruyter, 1975-2000.
- Scalise, S., Bisetto, A. (2009). « The classification of compounds ». In : R. Lieber, P. Štekauer (eds.), *The Oxford Handbook of Compounding*. Oxford : Oxford University Press, 34-53.
- Villoing, F. (2012). « French compounds ». *Probus*, 24/1, 29-60.