

Noms sous-spécifiés et constructions : analyse distributionnelle dans un corpus français diversifié en genres

Shell nouns and their constructions: A distributional analysis on a genre-diversified French corpus

Anaïs Vajnovszki

Université de Gand (GLIMS) & Université Sorbonne-Nouvelle (Lattice)

anaïs.vajnovszki@ugent.be & anaïs.vajnovszki@sorbonne-nouvelle.fr

Peter Lauwers

Université de Gand (GLIMS)

peter.lauwers@ugent.be

Dominique Legallois

Université Sorbonne-Nouvelle (Lattice)

dominique.legallois@sorbonne-nouvelle.fr

Résumé

Le but de cet article est d'analyser le comportement constructionnel de 16 noms sous-spécifiés en français. Les noms sous-spécifiés (Nss) sont des lexèmes abstraits, comme *problème* ou *idée*. Ils constituent une catégorie nominale dont l'une des particularités est de s'intégrer de manière préférentielle dans les constructions copulatives spécificationnelles, p. ex. *le problème est que* + complétive ou *la solution est de* + infinitif. Toutefois, d'autres mécanismes existent, tels que la construction paratactique (*Autre obstacle en matière d'enseignement : la formation des enseignants, des maîtres // la formation des enseignants constitue un obstacle*) ou la construction adnominale (*le problème du culte de l'empereur // le culte de l'empereur constitue le problème*). Nous pensons qu'une analyse outillée fondée sur le comportement constructionnel de ces noms pourrait permettre de répondre à plusieurs questions, notamment : (1) quels Nss sont concernés par un comportement distributionnel similaire en termes d'intégration dans ces constructions ? (2) ces catégories distributionnelles correspondent-elles à des types sémantiques de Nss ? (3) cette catégorisation révèle-t-elle l'existence d'un prototype ? (4) une synonymie entraîne-t-elle une appartenance à la même catégorie ? (5) existe-t-il des liens entre les constructions d'accueil ? Nous répondrons à ces questions en menant une analyse outillée sur un petit corpus diversifié en genres, composé de 16 Nss parmi les plus fréquents dans les constructions spécificationnelles, constructions considérées par de nombreux auteurs

comme un critère d'identification des Nss. Notre objectif est double : tout d'abord, regrouper les Nss en plusieurs classes sur la base de leur fréquence dans une gamme de constructions ; ensuite, analyser la composition de ces groupes de Nss pour en tirer des conclusions plus théoriques sur la structure interne de cette catégorie nominale et sur les éventuels mécanismes plus globaux à l'œuvre.

Mots-clés : Nss, constructions spécificationalles, grammaire des Constructions, analyse outillée, analyse de groupe

Abstract

The goal of this article is to analyze the constructional behavior of 16 French shell nouns. Shell nouns are nouns like *problem* or *idea*. They make up a nominal category whose main characteristic is its preference for specificational copular constructions, like *the problem is that* + conjunctive clause or *the solution is to* + infinitive. However shell nouns can enter other constructions, for example the paratactic construction (*Autre obstacle en matière d'enseignement : la formation des enseignants, des maîtres* 'Another obstacle in the field of teaching: teacher training' // *la formation des enseignants constitue un obstacle* 'teacher training is an obstacle') or the adnominal construction (*le problème du culte de l'empereur* 'the problem of the emperor's cult' // *le culte de l'empereur constitue le problème* 'the emperor's cult is the problem'). A bottom-up statistical analysis of a home-made corpus enables us to tackle the following questions: (1) which shell nouns display a similar distributional profile among the aforementioned constructions? (2) do these categories point to the existence of different shell noun types? (3) does this classification point to the existence of a prototype? (4) is synonymy linked to a similar distributional profile? (5) how are shell-noun-welcoming constructions related? We will answer these questions by analyzing the distribution of a small set of shell nouns which appear in a genre-diversified corpus and were selected as the 16 most frequent ones in the specificational construction, considered by many authors as criterion for shell-nounhood. Our goal is hence to shed light on the internal structure of this nominal category, as well as on the distribution of shell-noun constructions.

Keywords: shell nouns, specificational constructions, Construction grammar, corpus-based, cluster analysis

1. Introduction

Les noms sous-spécifiés (ci-après Nss) sont des noms abstraits (Schmid, 2000, par ex.) comme *problème* (1) ou *but* (2) :

- (1) **Le problème** c'est que tu as des élus qui veulent savoir comment ça marche (Corpus communications orales)
- (2) **Le but** est de réhabiliter le projet révolutionnaire. (Corpus presse)

Ces noms sont marqués par une incomplétude informationnelle qui leur permet d'« encapsuler » (Francis, 1986) une portion de texte appelée Contenu Propositionnel (CP). Ainsi en (1) *problème* est le Nss et la complétive son CP ; dans (2) c'est le Nss *but* qui est spécifié par son CP, encodé sous la forme d'un infinitif¹.

Les termes qui désignent ces noms sont multiples : *container nouns* (Vendler, 1968), *unspecific nouns* (Winter, 1992), *enumerables* (Tadros, 1985), *anaphoric nouns* (Francis, 1994), *carrier nouns* (Ivanič, 1991) ou encore *signalling nouns* (Flowerdew, 2002). Cet éclatement terminologique reflète des conceptions légèrement différentes : par exemple, les *anaphoric nouns* (ou *A-nouns*) de Francis (1994) ne peuvent fonctionner, comme leur nom l'indique, qu'en opérant un renvoi vers le contexte antérieur ; de même, selon Flowerdew (2002) pour l'anglais et selon Ho-Dac et al. (2022) pour le français, pour être en emploi Nss un nom doit forcément faire référence à une portion de texte, et pas simplement avoir « le potentiel » (Schmid, 2000) pour le faire.

Ces divergences parfois conceptuelles posent aussi la question du recoupement entre la catégorie des Nss et celle des noms généraux. Selon Halliday et Hasan (1976), les noms généraux, tels que *people*, *thing* ou *affair*, sont des « superordinates ». Les noms généraux ne sont donc pas tous des Nss, même si, selon Adler (2017) cette distinction est une affaire de degré.

Une autre caractéristique des Nss, qui est la conséquence de leur incomplétude informationnelle, est leur capacité à entrer dans des constructions permettant l'encapsulation, comme les constructions spécificationnelles (ou CS), exemplifiées en (1) et (2) *supra*, mais aussi la construction paratactique (3), dans laquelle c'est la ponctuation (à l'écrit) qui supplée l'absence de la copule (construction d'abord identifiée par Blanche-Benveniste, 1992), ou encore la construction résomptive (4). De nombreux auteurs ont décrit les constructions que privilégient les Nss (Blanche-Benveniste, 1992 ; Adler & Legallois, 2018 ; Ho-Dac et al., 2020), certains (Roze et al., 2014) faisant même de l'intégration d'un constituant dans une CS avec une complétive (1) ou un infinitif (2) un critère suffisant à l'identification d'un Nss.

¹ Pour des raisons de lisibilité, le Nss des occurrences illustrant cette étude est marqué en gras, et son CP est souligné.

- (3) **Le résultat** : tout a disparu. (Exemple tiré de Blanche-Benveniste, 1992)
- (4) Or un massif met au moins dix ans à retrouver son aspect initial. **Ce problème** est sujet à controverse aux antipodes car l’Australie a refusé de ratifier le protocole de Kyoto destiné à limiter le réchauffement de la planète, seul responsable du blanchiment des coraux. (Corpus presse)

Les Nss forment une catégorie ouverte (Schmid, 2000 par ex.), ce qui signifie que de nouveaux lexèmes, *a priori* moins adaptés à une utilisation en tant que « coquille conceptuelle » (5a), peuvent entrer dans cette catégorie, si tant est qu’ils réussissent à intégrer la construction. Pour apparaître dans des CS, un nom doit cependant satisfaire certaines contraintes sémantiques, comme le prérequis de l’abstraction, ce qui rend impossible un énoncé comme (5b).

- (5) a. **Ma religion** est de chercher la vérité dans la vie et pour la vie dans la vérité, sachant que je ne les trouve pas pendant que je vis. » (Wikipedia)
- b. ***Ma table** est de chercher la vérité dans la vie et pour la vie dans la vérité.

En revanche, tous les noms abstraits ne sont pas des Nss (Schmid, 2000). Par ailleurs, un lexème généralement utilisé en tant que Nss peut aussi avoir d’autres emplois, d’où notre choix du terme d’*emploi Nss*.

Pour identifier les emplois Nss, il existe plusieurs tests. Un premier consiste à poser la question « Quel est le N ?² » (Blanche-Benveniste, 1992) : si le N est bien un Nss, la réponse sera le CP. La reformulation de la phrase d’origine en une CS (test proposé par Riegel (1996) pour ce qu’il appelle les « noms à compléments propositionnels ») constitue une autre manière de déterminer si le Ns est en emploi Nss ou non : si celle-ci est possible, alors il s’agit bien d’un Nss. À ces deux tests, nous ajoutons la réversibilité en une prédicationnelle, avec la copule *être* ou tout autre verbe (semi-)copule (test utilisé par Markussen (2016) dans les structures du type [N1 de N2], pour déterminer si N2 donne un « fond constitutif » à N2).

Ainsi, *problème* en (6a) est un Nss, comme le prouvent les tests en (6b), (6c) et (6d). Mais le même nom en (7a) n’est pas en emploi Nss, comme le prouvent les paraphrases infructueuses en (7b), (7c) et (7d). Ici, *problème* forme plutôt, avec un complément post-nominal non-actualisé, un SN à forte cohésion interne, comme le prouve la paraphrase en (7e), inspirée d’un test fourni par Markussen (2016) pour identifier les relations hyponymiques dans les structures [N1 de N2].

² Riegel (1996) propose un test similaire pour les « noms à compléments propositionnels », à savoir les « questions-réponses spécifiantes » du type « Quel N ? – CP » ou encore « C’était quoi comme N ? – CP ».

- (6) a. En dehors **du problème** du culte de l'empereur, on peut relever trois formes principales de la propagande [...] (corpus académique)
- b. Quel est **le problème** ? Le culte de l'empereur.
- c. **Le problème**, c'est le culte de l'empereur.
- d. Le culte de l'empereur constitue **le problème**.
- (7) a. Mon père est en arrêt maladie car il y a des grave problème de dos [sic]. (Corpus forums en ligne)
- b. *Quel est le problème de mon père ? Le dos.
- c. #Son problème, c'est le dos³.
- d. */#Le dos constitue son problème.
- e. De quel type de problème s'agit-il ? D'un problème de dos. (Test inspiré de Markussen, 2016)

S'il paraît clair qu'une poignée de constructions accueille la plupart des emplois Nss, l'incidence de l'apparition d'un même nom dans ces constructions reste méconnue. En effet, une distribution comparable parmi plusieurs constructions pourrait révéler l'existence de groupes de Nss au fonctionnement similaire, montrant par-là la structuration interne de la catégorie. À cet effet, une étude outillée sur un petit échantillon de Nss parmi les plus fréquents dans certaines constructions discriminantes seulement pourrait apporter un éclairage différent à des analyses de grande envergure sur les Nss et leurs constructions, comme celles de Roze et al. (2014) ou Adler et Legallois (2018).

À partir de là, nous pourrions aborder plusieurs questions. Tout d'abord, il conviendra de se demander :

- 1) Quels Nss sont concernés par un comportement distributionnel similaire en termes d'intégration dans ces constructions ?

Cette analyse nous amènera à nous interroger sur l'existence de différents types de Nss :

- 2) Ces catégories purement distributionnelles correspondent-elles à des types sémantico-conceptuels de Nss ?

Ainsi que sur la question du Nss prototypique :

- 3) Cette catégorisation correspond-elle à l'existence d'un centre prototypique ?

³ Le signe # indique que cette phrase est acceptable mais qu'elle entraîne un changement de sens par rapport à (7a). En effet en (7c), *le dos* est envisagé par le locuteur comme un élément problématique, précisément dans ce contexte d'énonciation ; alors qu'en (7a), *de dos* (non-actualisé car sous sa forme « nue ») vient s'ajouter à *problèmes* pour créer un SN composé, sans pour autant y apporter un « fond constitutif » (termes de Markussen, 2016).

Nous répondrons à cette troisième question de recherche en sachant que l'échantillon examiné est constitué de 16 Nss⁴, qui comptent parmi les plus prototypiques, si l'on s'en tient au critère de la fréquence, déjà proposé par Schmid (2000). Par ailleurs, notons que le but de notre analyse n'est pas de découvrir en quoi consiste un Nss prototypique, mais plutôt de déterminer quels Ns correspondent à ce fonctionnement, en termes constructionnels.

Au passage, nous avons aussi vérifié si la synonymie entraîne un comportement distributionnel similaire, en répondant à la question suivante :

- 4) Une synonymie entraîne-t-elle une appartenance à la même catégorie ?

Enfin, si on part cette fois des constructions pour arriver aux Nss qui les intègrent, il est également pertinent de se poser la question suivante :

- 5) Existe-t-il des liens entre les constructions d'accueil ?

2. Les constructions identifiées par les études sur les Nss : aperçu et recoupements

Nous donnons *infra* la liste des constructions identifiées dans la littérature sur les Nss français, qui s'inspire largement des travaux de Schmid sur l'anglais (2000) :

⁴ La composition de cet échantillon de Nss a été décidée d'après plusieurs critères, parmi lesquels la fréquence du lemme dans certaines des CS identifiées comme étant les plus discriminantes par la littérature sur les Nss, ainsi qu'en prenant en compte d'éventuels effets de quasi-synonymie entre les lexèmes concernés, p. ex. *chose* et *truc* (voir Section 3.2 pour une description détaillée de la méthodologie).

Construction		Exemple
1.	CS avec un infinitif	La question est de <u>savoir ce que l'on prend comme élément de référence.</u> (Roze et al., 2014)
2.	CS avec une complétive	Le risque est que <u>ce genre de comportement rappelle de bien mauvais souvenirs.</u> (Roze et al., 2014)
3.	CS nominale	Le problème , c'est <u>le décalage entre vos intentions et la réalité.</u> (Adler & Legallois, 2018)
4.	Groupe nominal Nss expansé par un infinitif	C'est un site commercial qui a été lancé en juin 2002, à New York, avec l'objectif de <u>mettre les gens en relation les uns avec les autres autour d'un sujet d'intérêt commun.</u> (Roze et al., 2014)
5.	Groupe nominal Nss expansé par une complétive	L'idée que les inspecteurs puissent <u>renifler les armes et les documents qui s'y rapportent sans l'aide des autorités irakiennes</u> est absurde. (Roze et al., 2014)
6.	Groupe nominal anaphorique à l'initiale d'une phrase et déterminé par un démonstratif	Vous <u>sacrifiez la santé des Français les plus pauvres aux équilibres comptables.</u> Ce choix n'est pas le nôtre. (Adler & Legallois, 2018)
7.	Construction attributive caractérisante	L'argumentation du ministre est techniquement juste. <u>La situation des DOM et celle de la Corse sont profondément différentes.</u> (Adler & Legallois, 2018)
8.	Construction introduisant une spécification indépendamment de la syntaxe, à l'aide de « : »	Leur objectif : <u>être identifiés comme des « adultes disponibles » avec lesquels on peut parler de tout et de rien.</u> (Roze et al. 2014)

Construction		Exemple
9.	Construction avec spécification explicite par le verbe <i>consister à</i> ou <i>résider en</i>	[...] plusieurs journaux économiques ont de même souligné le paradoxe qui consiste à <u>ouvrir notre marché à la concurrence</u> . (Adler & Legallois, 2018)
10.	Construction impliquant un verbe support	Mon amendement avait pour objectif de <u>combler cette lacune</u> . (Adler & Legallois, 2018)
11.	Construction être {un / le} Nss dans laquelle le Nss catégorise ou évalue le sujet	<u>On préfère deux entreprises moyennes à une grande entreprise nationale</u> . C'est une erreur stratégique . (Adler & Legallois, 2018)

Tableau 1. Inventaire non-exhaustif des constructions d'accueil possibles des Nss (d'après Roze et al., 2014 et Adler & Legallois, 2018).

Selon nous, certaines des constructions *supra* peuvent être regroupées : par exemple, du point de vue constructionnel, la CS avec un infinitif (ligne 1 du Tableau 1) et celle avec une complétive (ligne 2) relèvent toutes les deux du mécanisme de la CS (seul le type de complément change). Nous considérons donc que les recoupements occasionnels de certaines constructions, pourtant présentées comme distinctes dans le Tableau 1 *supra*, sont inévitables, puisque les auteurs dont nous nous sommes inspirés pour établir ce tableau récapitulatif ont adopté une approche maximale précise, reconnaissant un nombre maximal de micro-constructions différentes. Nous avons quant à nous préféré regrouper les mécanismes à un niveau de granularité légèrement plus grossier, avec des catégories maximale différentes, possibilité offerte par la Grammaire des Constructions, qui reconnaît des constructions à différents niveaux de schématicité. Le tout est de bien déterminer à quel niveau d'abstraction nous opérons les distinctions entre ces constructions. Notre schéma d'annotation, qui adopte un autre découpage, sera présenté dans la Section 3.3.

3. Corpus et méthodologie

3.1. Corpus

Le corpus utilisé, étiqueté via l'étiqueteur automatique Tree Tagger de la plateforme Sketch Engine, contient 10 millions de mots⁵. Il a été constitué en combinant trois méthodes : 1) la collecte de textes au format texte (pour les corpus « académique » et « écrits de fiction ») ; 2) le recours à des corpus sources, comme *Le Monde* (2003-2004) (corpus « articles de presse ») ou le CEFC (corpus « communications orales ») ; 3) le moissonnage de plusieurs sites Web, notamment le site de l'Assemblée nationale et divers forums en ligne, directement via la plateforme Sketch Engine (pour les corpus « discours & allocutions » et « forums en ligne »). Le corpus global est divisé en 6 sous-corpus de taille comparable, et correspondant à 6 genres textuels.

Corpus	Taille (en nbre de mots)
Académique	1 896 918
Écrits de fiction	1 683 848
Articles de presse	1 754 442
Forums en ligne	1 971 605
Discours & allocutions	1 779 772
Communications orales	1 291 666
Total	10 378 251

Tableau 2. Composition du corpus.

De nombreux auteurs soulignent l'importance d'une étude contrastive du point de vue du genre textuel (Schmid, 2000 ; Flowerdew & Forest, 2014 ; Ho-Dac *et al.*, 2020). En effet, la plupart des études sur corpus sur les Nss en français ne portent que sur un genre textuel, à savoir le genre article journalistique (Roze *et al.*, 2014) et les allocutions politiques (Adler & Legallois, 2018), qui plus est des genres marqués par de la non-spontanéité. Or, comme les Nss servent une fonction d'organisation discursive (Legallois & Gréa, 2006 ; Lefevre, 2007), nous pensons qu'exploiter un corpus diversifié en genres permettra d'identifier des fonctionnements inédits.

⁵ Corpus constitué dans le cadre d'une thèse de doctorat (en cours), supervisée par D. Legallois et G. Cislaru (Paris 3), et par P. Lauwers, M. Van Peteghem et P. Hadermann (Université de Gand).

3.2. Méthodologie

Pour dresser une liste de Nss potentiels, nous nous sommes inspirés de l'analyse sur structures syntaxiques spécifiques de Roze et al. (2014), qui ont procédé en deux temps :

- Une extraction des noms entrant dans les structures *N est de Inf.* (ligne 1 du Tableau 1) et *N est que P* (ligne 2) ;
- Une projection en corpus des quelque 300 lemmes identifiés suite à la première étape de l'analyse, afin d'identifier l'ensemble des structures auxquelles ces Nss (potentiels) participent.

3.2.1. Extraction et projection des lemmes en corpus

Dans notre cas, la phase d'extraction des noms nous a permis d'identifier près de 300 lemmes nominaux apparaissant à différentes fréquences dans les structures ciblées, totalisant plus de 3 000 occurrences. Par exemple, le lemme *problème* apparaît 292 fois dans ces structures dans l'ensemble du corpus (voir Tableau 3).

La deuxième phase de projection a relevé toutes les occurrences de chacun de ces noms dans l'ensemble du corpus, peu importe la construction dans laquelle le nom apparaît. Ainsi, le mot *problème* a une fréquence absolue de 6 428 (voir Tableau 3).

L'attraction de ces noms pour les constructions ciblées par la première étape a ensuite été calculée, en divisant la fréquence d'apparition du nom dans ces structures par sa fréquence totale dans le corpus. À titre d'exemple, *problème* (1^{ère} ligne du Tableau 3) a une attraction de 4 % pour les constructions de la première étape, alors que l'attraction de *chose* (5^{ème} ligne) pour ces mêmes constructions n'est que de 1,17 %.

Lemme	Fréquence absolue	Fréquence dans les structures de la phase d'extraction	Attrance du lemme pour les structures de la phase d'extraction
problème	6 428	292	4 %
objectif	1 919	275	14,33 %
but	5 017	199	3,96 %
idée	3 465	128	3,69 %
chose	8 490	100	1,17 %

Tableau 3. Les 5 premières lignes des calculs permettant de déterminer l'attrance, pour les structures ciblées par la phase d'extraction, des 300 lemmes identifiés lors de cette première phase d'extraction.

3.2.2. Sélection de 16 Nss parmi les plus représentatifs

Pour des raisons de faisabilité, parmi cette liste d'environ 300 Nss avérés pour lesquels l'attraction a été calculée, nous n'en avons retenu que 16 parmi les plus fréquents. Cet échantillon a été sélectionné selon des critères, entre autres : (i) de fréquence dans les structures ciblées par la phase d'extraction (colonne 3 du Tableau 3) ; (ii) d'attrance pour ces structures (colonne 4 du Tableau 3) ; et (iii) d'éventuels effets de quasi-synonymie (à condition que le synonyme apparaisse au moins 100 fois dans l'ensemble du corpus) : ainsi, *problème*, *difficulté* et *obstacle* ont tous les trois été retenus car leur analyse nous permettra de savoir si une proximité sémantique correspond forcément à un comportement distributionnel analogue, en termes de constructions (question de recherche 3) ; à l'inverse, *souci*, *hic* et *point noir* n'ont pas été retenus en raison de leur trop faible fréquence dans le corpus. Cet échantillon se compose ainsi des 16 noms suivants : *but*, *chose*, *conviction*, *devoir*, *difficulté*, *idée*, *mission*, *objectif*, *obstacle*, *problème*, *question*, *rôle*, *solution*, *tâche*, *truc*, *vérité*.

3.2.3. Annotation et échantillonnage des 100 premières occurrences aléatorisées

Pour constituer le corpus correspondant à cet échantillon de 16 noms, un long processus d'annotation a été mené, au cours duquel nous avons identifié à chaque fois la construction d'accueil. Concrètement, pour chacun des noms retenus, nous avons aléatorisé les occurrences et annoté les 100 premières occurrences correspondant à des emplois abstraits, ce qui donne un corpus initial d'une taille de 1 600 occurrences. Au cours de cette phase d'annotation détaillée, nous avons évacué les exemples qui ne correspondent pas, entre autres, à des emplois non-abstraites, comme (8) *infra*, où le nom collectif *mission* dans le SN *la mission d'information* désigne, par métonymie, un « groupe de personnes ayant une mission » (*Le Petit Robert*) :

- (8) Je pense que nous allons respecter cette décision très claire de la mission d'information : j'émet un avis défavorable sur votre amendement. (Corpus discours & allocutions)

Le schéma d'annotation utilisé lors de cette étape du travail est inspiré du Tableau 1 *supra*, à partir duquel nous avons réalisé certains regroupements (mais aussi certains détachements) des structures identifiées par les auteurs précédents. En effet, certaines constructions présentées dans le Tableau 1 sont des variantes qui, à un niveau plus schématique (d'après la terminologie de la linguistique cognitive, p. ex. Langacker, 1987), relèvent d'un mécanisme plus global (tandis que d'autres relèvent de mécanismes ayant été érigés au rang de construction à part entière, p. ex. la locution prépositive). Ces regroupements (et éclatements) sont résumés dans le Tableau 4 :

Construction telle qu'identifiée par les études existantes (cf. Tableau 1 <i>supra</i>)		Notation constructionnelle
1.	CS avec un infinitif	[Nss copule (/assimilé) CP]
2.	CS avec une complétive	
3.	CS nominale	
9.	Construction avec spécification explicite par le verbe <i>consister à</i> ou <i>résider en</i>	
11.	Construction être {un / le} Nss dans laquelle le Nss catégorise ou évalue le sujet	
4.	Groupe nominal expansé par un infinitif ⁶	[Nss <i>de</i> (/assimilé) CP]
5.	Groupe nominal Nss expansé par une complétive ⁷	
6.	Groupe nominal anaphorique à l'initiale d'une phrase et déterminé par un démonstratif	[CP </s ⁸ > dét. NSS] [NSS </s> CP]
7.	Construction attributive caractérisante	
8.	Construction introduisant une spécification indépendamment de la syntaxe, à l'aide de « : » ⁹	[NSS ponct. CP (ponct.)] [CP ponct. Nss (ponct.)]
10.	Construction impliquant un verbe support ¹⁰	[<i>avoir</i> (/assimilé) prép. Nss <i>de</i> CP]
4.	Groupe nominal expansé par un infinitif ¹¹	[<i>dans</i> dét. Nss <i>de</i> CP]

Tableau 4. Les 11 constructions du Tableau 1 regroupées selon le mécanisme encapsulant.

À ce niveau de granularité, chaque construction retenue dans le schéma d'annotation final (présenté dans le Tableau 5 *infra*) s'oppose maximalement aux autres macro-constructions. Par ailleurs, cela n'exclut pas que ces constructions plus spécifiques puissent être distinctives pour tel ou tel Nss, mais il a fallu sélectionner les paramètres maximalement différents en tenant compte du niveau d'abstraction de la taxonomie, en l'occurrence celui des stratégies formelles d'encapsulation majeures. Ce choix réfléchi résout aussi d'éventuels problèmes liés à la pénurie de données, c'est-à-dire des catégories (ici des constructions) trop peu peuplées pour mener un traitement quantitatif.

⁶ Traitement appliqué uniquement aux occurrences où le SN n'est *pas* précédé d'une préposition. Les occurrences précédées d'une préposition ont été classées comme appartenant à la construction prépositive [*dans* dét. Nss *de* CP] (dernière ligne du Tableau 4).

⁷ *Idem.*

⁸ <s> est un signe de ponctuation non-encapsulant ou une frontière de phrase.

⁹ Traitement appliqué uniquement aux occurrences où la ponctuation joue seule le rôle de la copule (i.e., pas de copule présente).

¹⁰ Traitement appliqué uniquement aux occurrences où le CP constitue l'attribut de l'objet du Nss.

¹¹ Traitement appliqué uniquement aux occurrences où le SN est précédé d'une préposition.

Notre schéma d'annotation, basé sur les regroupements (et éventuels éclatements) présentés dans le Tableau 3, est quant à lui présenté ci-dessous :

	Construction	Annotation constructionnelle et description	Exemple
1.	Emploi absolu	[Nss] Le locuteur ne fournit pas de CP.	Il faut avant de partir bien avoir en tête le but de son voyage et les attentes quant à son projet personnel [...]
2.	CS	[Nss copule (/assimilé) CP] Une construction copulative spécificationnelle qui permet une encapsulation de manière cataphorique.	Le problème c'est que <u>tu as des élus qui veulent savoir comment ça marche</u>
3.	Copulative prédicationnelle	[CP copule (/assimilé) Nss] Un verbe (semi -) copule permet au Nss d'encapsuler le CP de manière anaphorique (pointage vers la gauche).	<u>Demander l'aide d'une personne extérieure</u> est en effet l'une des solutions à envisager , pour toi et ta mère.
4.	Construction illustrative	[Nss <i>comme</i> (/assimilé) CP] Le CP n'est que partiellement donné.	Ce sont des obstacles comme celui de <u>l'affiliation à l'assurance-maladie</u> qui donnent à nos compatriotes l'impression d'être discriminés.
5.	Emploi adnominal	[Nss <i>de</i> (/assimilé) CP] Le CP est encodé sous la forme d'un complément du Nss.	Du moins l'opinion, surtout parmi leur personnel, avait-elle la conviction qu'il en était ainsi .
6.	Cohésion transphrastique	[CP </s> dét. NSS] [NSS </s> CP] Le CP se trouve dans une autre phrase que le Nss.	Il s'agit pour eux, a-t-il déclaré à la BBC, de <u>certifier si Saddam coopère ou non</u> . Ce devoir ne veut pas dire seulement qu'il doit permettre aux inspecteurs d'accéder aux sites.
7.	Parataxe	[NSS ponct. CP (ponct.)] [CP ponct. Nss (ponct.)] Le Nss et son CP sont coordonnés via une parataxe : ponctuation, blanc ou conjonction de coordination.	[...] les difficultés commencent durant la troisième : <u>les cravings surviennent sans raison, au lever, le matin</u> .

	Construction	Annotation constructionnelle et description	Exemple
8.	Structure à Attribut de l'Objet	[avoir (/assimilé) prép. Nss de CP] Le locuteur assigne un CP au Nss grâce à une construction attributive de type attribut de l'objet (prépositionnel).	[...] lorsqu'on n'admet pour seule mission à la sociologie que <u>d'expliquer la vérité du réel</u> [...]
9.	Locution prépositive	[dans dét. Nss de CP] Le Nss et son CP adnominal forment une locution introduite par une préposition.	Vous avez retenu les dispositions minimales de ce texte dans le seul objectif de pouvoir afficher une loi .

Tableau 5. Les 9 constructions syntaxiques retenues dans notre schéma d'annotation morpho-syntaxique, inspirées des constructions citées dans le Tableau 1.

Précisons que l'échantillon initial répond à une politique d'inclusion maximale, ce qui signifie que nous avons inclus tous les emplois dans cet échantillon, y compris les emplois absolus comme en (10), où l'existence du référent CP est niée, et les emplois correspondant à l'exemple (7) présenté en introduction.

En effet, nous ne voyons pas le renvoi vers une portion de texte, parfois plus ou moins vague, comme une propriété nécessaire des Nss (contrairement à Flowerdew & Forest, 2014 ou Ho-Dac et al., 2022). Nous considérons donc les Nss « vides », c'est-à-dire ceux dont le CP n'est pas récupérable en contexte, comme des Nss à part entière. Nous nous basons pour cela sur la définition originale de Schmid (2000) : les « shell nouns » ont « le potentiel » nécessaire pour encapsuler une portion de texte, mais la matérialisation de ce pouvoir encapsulateur ne doit pas nécessairement prendre la forme d'une portion de texte directement identifiable. On peut ici poser une analogie avec l'emploi absolu d'un verbe, qui lui aussi correspond à la non-saturation d'un complément essentiel. Selon le cas, l'omission d'un CP peut être due, comme dans le cas des verbes en emploi absolu, à la possible récupération de celui-ci d'après la situation (9) ou à une volonté de généralisation (10).

- (9) En effet, le dimanche non chômé n'est, dans notre culture, que le fait de personnes exerçant **des devoirs envers la société**. Forces de l'ordre, médecin, personnels hospitaliers, personnes valorisant notre patrimoine et notre civilisation, tous ceux-là sacrifient de leur temps parce qu'ils relèvent d'une obligation sociale, d'un sens du devoir, d'une obligation morale. (Corpus discours & allocutions)
- (10) Si nous tenons à saluer le long et patient travail conduit par notre collègue Laurent Grandguillaume, il faut nous rendre à l'évidence : ce texte se heurte à une problématique très complexe et ne peut prétendre, à lui seul, résoudre **les difficultés grandissantes rencontrées par les chauffeurs**, quelle que soit leur

situation professionnelle – chauffeurs de taxi ou chauffeurs de VTC. (Corpus discours & allocutions)

Nous avons donc inclus des cas correspondant à (11a), où l'ajout d'une spécification (11b) est acceptable. Pour déterminer s'il s'agissait bien d'un Nss en emploi absolu, le test suivant a été appliqué : si l'ajout d'une spécification sous la forme d'une relative appositive ou un autre procédé est possible, alors l'exemple en question est considéré comme un Nss, certes vide mais en emploi sous-spécifié (une « coquille conceptuelle », selon les termes de Schmid, 2000), et annoté comme relevant de l'emploi absolu. Parfois, lorsque l'exemple original nie l'existence du référent du Nss (12a), le test inclut également la réfutation de cette négation (12b) dans une structure exceptive. Notons que notre corpus contient bien des exemples originaux correspondant à (12b), où l'existence du Nss est dans un premier temps niée pour que son CP soit ensuite mieux prédiqué dans le cadre d'une restriction (13).

- (11) a. Il faut avant de partir bien avoir en tête **le but** de son voyage et les attentes quant à son projet personnel, participer activement sur place tout comme se relaxer sont possibles à Moidjio. (Corpus forums en ligne)
- b. Il faut avant de partir bien avoir un tête **le but** de son voyage, qui est de s'amuser.
- (12) a. Nous n'avons eu **AUCUN problème** durant notre séjour de 7 mois. (Corpus forums en ligne)
- b. Nous n'avons eu **AUCUN problème** durant notre séjour de 7 mois, si ce n'est que les voisins étaient un peu bruyants.
- (13) Je n'ai **aucune idée précise** des dates si ce n'est que ça ne sera pas avant janvier ou février 2008. (Corpus forums en ligne)

3.2.4. Difficultés liées à l'annotation

Quelques occurrences dans lesquelles le Nss était séparé de son CP ont présenté une difficulté supplémentaire quant à l'identification de la construction d'accueil : en effet, l'exemple (14) correspond-il (i) à un cas de cohésion transphrastique (fonctionnement d'abord mentionné par Winter (1982), qui parle de « signalling nouns » qui encapsulent « across the boundaries of sentences » ; voir ligne 4 du Tableau 3 supra), dans lequel le point marque une frontière entre la proposition contenant le Nss et celle contenant son CP ? Ou s'agit-il (ii) d'une construction paratactique (voir ligne 5 du Tableau 3) dans laquelle la ponctuation n'est pas la ponctuation canonique (i.e., les deux points), mais le point final ?

- (14) La position que viennent de prendre la rapporteure et la secrétaire d'État soulèvent **des difficultés très sensibles** dans plusieurs régions. Ainsi, le découpage que vous avez imposé aux Alsaciens fait qu'ils n'auront pas la liberté d'engager ce type de conventions. Il en est de même, dans le monde basque, comme notre collègue Capdevielle vient de le souligner. Et c'est un vrai souci en Loire-Atlantique, où l'on note un développement de l'enseignement du breton. (Corpus discours & allocutions)

Nous avons tranché ce problème en appliquant le test suivant : si le point final se laisse remplacer par les deux points, comme en (14), alors la portion de texte qui suit la ponctuation est le CP du Nss, indiquant une structure paratactique.

En outre, l'identification d'un seul mécanisme encapsulant a posé problème lorsqu'un Nss mobilise plusieurs des constructions mentionnées dans le Tableau 3. C'est par exemple le cas en (15), qui participe à la fois au cas de figure décrit dans la ligne 2 (encapsulation cataphorique via un verbe (semi-)copule), mais aussi dans la ligne 3 (encapsulation anaphorique via un verbe (semi-)copule) et dans la ligne 7 (encapsulation paratactique, via la ponctuation) :

- (15) a. Mais la « liberté de marché », poursuit-il, n'a de sens qu'à des conditions de concurrence bien moins inégales qu'aujourd'hui. C'est **sa deuxième conviction** : l'échange n'est pas libre quand il met aux prises un Nord qui a tous les attributs de la puissance et un Sud qui en est largement dépourvu. (Corpus presse)
- b. **Sa deuxième conviction** est que la liberté de marché n'a de sens qu'à des conditions de concurrence bien moins inégales qu'aujourd'hui.
- c. **Sa deuxième conviction** est que l'échange n'est pas libre quand il met aux prises un Nord qui a tous les attributs de la puissance et un Sud qui en est largement dépourvu.

Dans ce cas précis, étant donné l'ambiguïté référentielle du pronom *ce*, qui peut pointer aussi bien vers la gauche que vers la droite, nous avons penché pour l'interprétation selon laquelle le Nss encapsule le CP qui suit la copule (cataphore), en raison du lien plus exclusif et donc plus saillant que la copule établit entre le Nss et son CP. Il reste toutefois vrai que cette réduction à un seul mécanisme encapsulant ne rend pas compte de toute la complexité référentielle des Nss.

3.2.5. Nettoyage pour la présente étude

Dans le cadre de la présente étude, nous avons écarté du corpus initial de 1 600 occurrences les exemples qui ne correspondent pas à des emplois Nss, comme en (7) dans l'introduction. Nous considérons que ces emplois adnominaux constituent un emploi catégorisant,

taxonomique, dans lesquels la relation entre le premier N et le deuxième est une « relation hyponymique [...] d'inclusion catégorielle » (Markussen, 2016, p. 141), c'est-à-dire dans lesquels est prédiquée l'existence d'un sous-type : en l'occurrence, un type de problème. Syntactiquement, le N après la préposition *de* n'est pas précédé de l'article (emploi non-actualisé de *dos* en (7)) ; sémantiquement, il ne saurait constituer le CP du Nss dans une prédicationnelle / une CS.

Le reste des emplois a été inclus à notre analyse, quitte à ratisser plus large que la plupart des auteurs et à inclure des occurrences qu'un certain nombre d'auteurs ne qualifieraient pas d'emplois Nss, dans la mesure où ils ne sont pas saturés par un CP (emploi qui, comme nous le verrons, s'est même avéré constitutif d'un groupe de Nss).

Après ce nettoyage final, le corpus réduit pour notre analyse comprend désormais 1 426 occurrences. Par rapport au corpus de départ (1 600 occurrences), ce corpus réduit présente parfois une légère asymétrie selon le lemme analysé : ainsi, l'échantillon final pour *vérité* contient 41 occurrences, mais celui pour *but* en contient 100. Cette différence n'entravera toutefois pas la représentativité de nos résultats car les conclusions de l'analyse de groupe sont basées sur des quantités relatives, dont le total, par variable, s'élève à 100 (voir infra). Seulement, les proportions de Nss qui comptent 100 occurrences sont un peu plus robustes que celles des Nss qui en comptent moins.

3.3. Outils statistiques

Afin de répondre à nos questions de recherche, nous avons mobilisé plusieurs techniques statistiques. Nous avons réalisé d'abord une analyse de groupe (*cluster analysis*) basée sur le profil distributionnel de nos Nss, le but étant de les regrouper d'après leur apparition dans les constructions d'accueil. À cet effet, nous avons mené une *Behavior Profile Analysis* (BPA) à l'aide du script de Gries (2008). Cette méthode a fourni les données brutes pour notre analyse, à savoir un tableau composé de 16 vecteurs (i.e., 16 lignes) contenant la fréquence relative du Nss dans chacune des constructions (i.e., les colonnes), le total étant égal à 1 pour chaque Nss. Les distances entre ces vecteurs, comprises entre 0 et 1 (*chose* et *difficulté* étant les voisins les plus proches, par exemple, à une distance euclidienne de 0,09), ont ensuite servi de base à une analyse de groupe hiérarchique (réalisée avec la fonction *hclust* du package Stats (R core)), selon la méthode de Ward. Il en résulte un dendrogramme qui visualise les regroupements des Nss¹² (voir Figure 1, Section 4), les Nss au comportement constructionnel similaire étant regroupés d'abord.

Pour mettre au jour les constructions majeures responsables des regroupements obtenus, nous avons complété nos résultats de deux manières. Tout d'abord, une analyse bivariée

¹² La qualité du clustering est bonne, comme le montre le coefficient agglomératif (0,79 ; valeurs entre 0 et 1).

croisant les Nss et les constructions nous a permis de voir quelles constructions sont sur- ou sous-représentées pour un Nss donné. À cet effet, nous avons calculé un khi deux et examiné les résiduels standardisés (i.e., l'écart par rapport aux fréquences attendues) pour les visualiser ensuite à l'aide de graphiques en mosaïque générés par le R package *vcd* (Meyer et al., 2023). (voir Figure 2). Cette analyse bivariée permet donc de voir quelles constructions sont attirées (bleu) ou repoussées (rouge) par tel Nss, mais ne dépasse pas encore le niveau des *Nss individuels*.

Pour voir quelles constructions sont attirées ou repoussées au niveau des *groupes* de Nss cette fois, nous avons calculé la moyenne par groupe des pourcentages des vecteurs de la BPA. Nous en avons ensuite tiré un graphique serpentin (« snake plot », Gries & Divjak, 2009, d'après le code fourni par Levshina, 2015, pp. 313-314) qui visualise les constructions maximale-ment distinctives, responsables des oppositions entre les groupes. Toutes ces techniques nous ont permis de découvrir des regroupements de Nss, ainsi que les constructions qui les sous-tendent. Enfin, afin de nous faire une idée des relations entre toutes ces constructions, nous avons mené une analyse en composantes principales (Kassambara, 2017 ; R-packages FACTOMINER (Lê et al., 2008) et FACTOEXTRA (Kassambara & Mundt, 2020)). Cette technique regroupe les variables (en l'occurrence les constructions) autour de deux dimensions plus abstraites et maximale-ment différentes (appelées composantes principales), permettant de voir quelles constructions apparaissent le plus souvent ensemble par rapport à tous les Nss.

4. Résultats de l'analyse de groupe

L'analyse de groupe selon la méthode Ward révèle 4 groupes (voir Figure 1), dont les trois premiers forment un macro-groupe (en vert) :

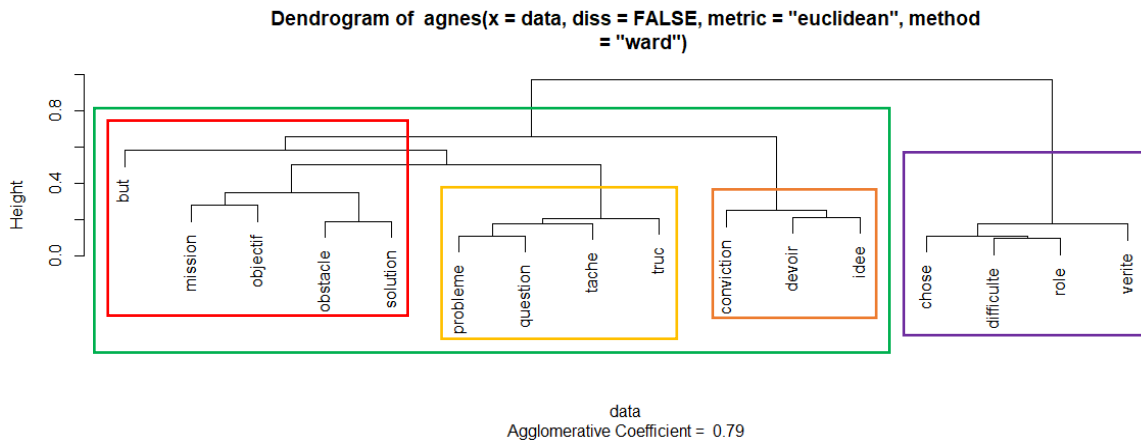


Figure 1. Les analyses de groupe ont identifié 4 groupes principaux parmi les Nss de notre échantillon, dont les 3 premiers forment un macro-groupe (encadrement vert).

Précisons que nous avons délimité l'analyse à partir de la hauteur 0,4 dans le graphique représenté dans la Figure 1. Concrètement, cela signifie que nous avons pu opposer les groupes 1 (en rouge) et 2 (en jaune), tout en sachant, par ailleurs, qu'ils appartiennent à un macro-groupe commun. Précisons également que *but* (à l'extrême gauche) se comporte comme un « outlier » en raison de son intégration privilégiée dans la locution prépositive *dans le but de*, dont il est particulièrement friand (voir Section 4.2.1 pour plus d'explications sur le comportement à part de *but*). Toutefois, en raison de sa préférence pour d'autres constructions constitutives du groupe 1 (en rouge), il sera ici traité comme un membre, certes éloigné, de ce groupe (voir Section 4.1 pour une analyse approfondie du groupe 1).

À partir de là, il conviendra de se demander pourquoi tel ou tel Nss a été réuni dans un même groupe, en complétant l'analyse de groupe par des analyses bivariées et des *snake plots*. Le premier constat qui se dégage est que le groupe 4 (violet) est isolé des 3 autres. Cette première bipartition entre le groupe 4 et le macro-groupe contenant les groupes 1, 2 et 3 nous pousse à commenter la composition du groupe 4 en premier.

4.1. L'opposition entre le groupe 4 et les 3 autres groupes : une question d'emploi absolu ?

Le groupe 4, un groupe à part, semble contenir les Nss en marge du spectre d'emploi sous-spécifié. Pour rappel, ce groupe contient les Nss suivants : *chose*, *difficulté*, *rôle* et *vérité*.

Au niveau des macro-groupes, on voit qu'ils s'organisent surtout en termes d'opposition entre, d'une part, les Nss qui sont le plus souvent remplis en contexte (en vert) et ceux qui apparaissent la plupart du temps en emploi absolu (en violet).

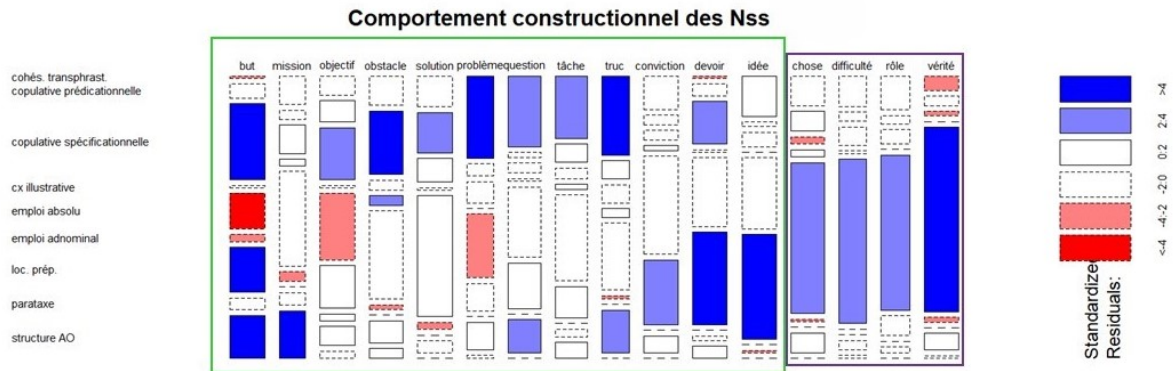


Figure 2. Résultats de l'analyse bivariée¹³ montrant le comportement constructionnel des Nss de notre échantillon¹⁴.

La Figure 2 montre quels Nss du groupe 4 (encadrement violet) sont caractérisés par une surreprésentation de l'emploi absolu (les barres en bleu et bleu foncé dans le groupe à droite), alors que les Nss dans cet emploi sont souvent sous-représentés (les barres rouges) dans le macro-groupe à gauche (encadrement vert). Cette interprétation est confirmée par la *snake plot*¹⁵ de la Figure 3 ci-dessous, qui montre des tendances similaires au niveau des groupes.

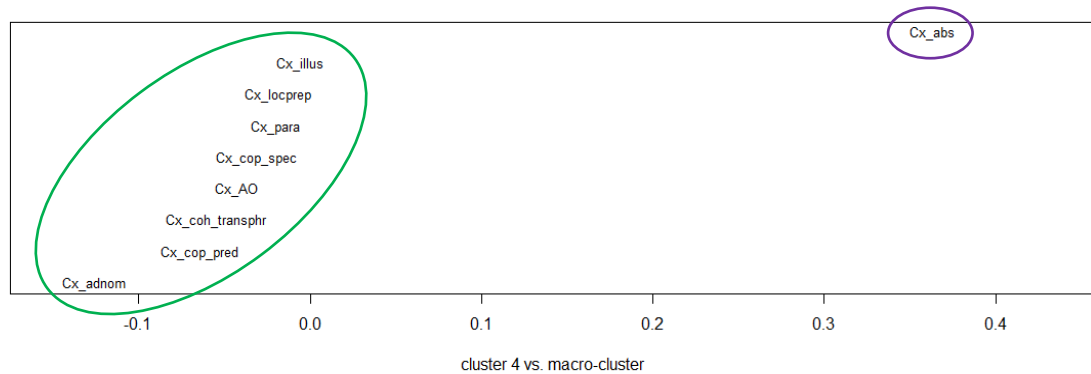


Figure 3. Snake plot montrant que la construction la plus clivante derrière l'opposition entre le macro-groupe (en vert) et le groupe 4 (en violet) est bien l'emploi absolu (en haut à droite).

¹³ La taille des boîtes représente le nombre d'observations.

¹⁴ Étiquettes raccourcies utilisées dans les graphiques :

Cohés. transphrastique = Construction à cohésion transphrastique

Cx illustrative = Construction illustrative

Loc. prép. = Locution prépositive

Structure AO = Construction à attribut de l'objet

¹⁵ Ou « graphique serpentin » (Van Wettere, 2018).

Le graphique serpentin de la Figure 3 représente les variables les plus importantes derrière l'opposition entre deux groupes, en l'occurrence : d'une part, les Nss qui forment le groupe 4 (les constructions positivement responsables de l'opposition sont entourées en violet) ; d'autre part, le reste des Nss (les constructions négativement responsables de cette opposition sont entourées en vert). La Figure 3 permet donc de contraster les variables qui « font la différence » pour distinguer un groupe donné par rapport aux autres groupes

Si la capacité à encapsuler un CP est l'une des caractéristiques principales des Nss, quelles propriétés sémantiques inhérentes aux Nss du groupe 4 (*chose*, *difficulté*, *rôle* et *vérité*) leur permettent alors d'échapper à des contextes encapsulants ?

4.1.1. Des noms abstraits en emploi générique

Étant donné que le corpus académique contient des textes issus d'ouvrages philosophiques, une grande majorité des occurrences du nom abstrait *vérité* correspond à un terme métaphysique. Ainsi, en (16), *vérité* est utilisé en tant que nom en emploi générique, pris dans son extension maximale, et n'a donc plus besoin d'être spécifié.

- (16) a. Selon la métaphysique, **la vérité** améliore et rend heureux. (Corpus académique)

Les exemples de ce genre ont tout de même été inclus dans notre corpus car, même si elle change l'extension du nom *vérité*, l'ajout d'une spécification est bien possible (16b) et même attesté en corpus (17) :

- (16) b. Selon la métaphysique, **la vérité selon laquelle nul n'est fondamentalement méchant** améliore et rend heureux.

- (17) **La vérité selon laquelle**, absolument parlant, tout est faux. (Corpus académique)

4.1.2. *Chose* utilisé comme un quasi-pronom auto-saturé

Parmi ces noms généraux, nous retrouvons aussi le « nom caméléon » (Kleiber, 1987) *chose*, un nom dont l'extension est « *a priori* maximale » (Benninger, 2014). Cette « polyvalence référentielle » est telle qu'il peut dénoter des « éléments non identifiés » (Huyghe, 2015). Grâce à cette « pauvreté descriptive » (Huyghe, 2015), *chose* peut être utilisé en tant que « quasi-pronom » (terme présent dans Halliday & Hasan, 1976, dans Francis, 1986 ; la comparaison entre *chose* et des mots comme *ça*, *cela* et *ce* est également présente chez Kleiber, 1987). Lorsque c'est le cas, on peut remplacer le SN par *quelque chose* :

- (18) de nombreuses choses ont évolué dans ce concept depuis la parution de ses avis datant de 2017 (Corpus forums en ligne)

Ces emplois, certes abstraits mais pris dans leur extension maximale, concernent des lexèmes qui, dans d'autres contextes (19), se comportent comme des Nss plus prototypiques, mais qui se retrouvent ici dans des emplois manifestement moins Nss. Toujours marqués par cette incomplétude informationnelle, les Nss utilisés ainsi sont certes toujours « en attente d'un contexte » (Ivanič, 1991), mais comme ils sont auto-saturés, s'appliquant à tous les CP potentiels, ce contexte ne vient jamais et l'encapsulation n'est donc pas réalisée. Or, comme le travail de délimitation préalable prenait en compte tous les emplois pour lesquels l'ajout d'une spécification était possible, ces d'emplois plus marginaux se retrouvent dans notre corpus.

- (19) Bonjour, je n'ai qu'**une chose à vous conseiller** c'est d'écouter votre cœur car vous avez une vie à vivre et votre fils aussi. (Corpus forums en ligne)

Notre politique d'inclusion maximale, ambitionnant une vue globale du phénomène, a certes conduit au traitement d'emplois plus marginaux, mais il n'était pas donné d'avance que ces emplois seraient constitutifs de tout un groupe de Nss : il ne s'agit donc pas d'emplois isolés, marginaux au sein de chaque Nss, ou caractérisant tous les Nss au même degré. L'émergence de tout un groupe est ainsi un résultat inattendu, conséquence directe d'une volonté de ne pas trop délimiter le champ à l'avance.

4.1.3. L'emploi comme pluriel lexical de *chose* et *difficulté*

Enfin, il convient aussi de mentionner l'emploi de type « pluriel lexical » (Lauwers, 2021) de *choses* (et, dans une moindre mesure, de *difficultés*). Tout comme d'autres noms au pluriel de ce type (*[partir dans] les montagnes, les environs, les intempéries*), le pluriel tend à se détacher du singulier, ce qui bloque la dénombrabilité (voir Lauwers, 2021, Section 2.2) : **/#Les trois choses ont changé depuis*). En ce qui concerne leur degré de sous-spécification, ou d'appartenance à la classe des Nss (« shell-nounhood », selon Schmid, 2000), *choses* et *difficultés*, mis au pluriel, subissent donc un changement sémantique, qui les rend moins aptes à fonctionner dans certaines constructions des Nss, notamment la CS, comme le montre les paraphrases impossibles en (20b) et (21b). De ce fait, le pluriel les rend « moins Nss », mais ils partagent néanmoins encore avec les Nss plus prototypiques, comme les emplois paratactiques encapsulant via les deux points, exemplifiés en (20) et (21).

- (20) a. Au-delà des arguments juridiques, **je veux dire les choses très clairement** : une telle dérogation freinerait l'accès des petites et moyennes entreprises aux marchés publics. (Corpus discours & allocutions)
- b. ***les choses** sont qu'une telle dérogation freinerait l'accès des petites et des moyennes entreprises aux marchés publics

- (21) a. la première semaine, on est souvent anxieux et déprimé, les deux suivantes sur un petit nuage et **les difficultés** commencent durant la troisième : les cravings surviennent sans raison, au lever, le matin. (Corpus forums en ligne)
- b. ***les difficultés** sont que les cravings surviennent sans raison, au lever, le matin.

4.1.4. L'emploi caractérisé de rôle

Une autre utilisation propre aux Nss du groupe 4 contribue à les éloigner du centre prototypique de la catégorie des Nss : il s'agit d'un emploi Nss accompagné d'une caractérisation, dans lequel la pertinence (22) ou, plus rarement, la non-pertinence (23) du SN, est appréciée grâce à son insertion dans la locution verbale *jouer le rôle* de (22) et/ou d'adjectifs (22-23) :

- (22) Nous déplorons également la baisse de 60 millions d'euros de la taxe affectée aux CCI **qui jouent pourtant un rôle clé** pour accompagner nos entreprises et leur permettre de se développer. (Corpus discours & allocutions)
- (23) [...] bien que l'hypothèse des extraterrestres créateurs de l'homme et dispensateurs de la civilisation **jouât un rôle plutôt secondaire** dans *Le matin des magiciens*. (Corpus académique)

Par ailleurs, même lorsque ces mécanismes ne sont pas mis en place, le SN dans un emploi caractérisé se laisse paraphraser par une structure contenant un adjectif exprimant l'importance, p. ex. *à quel point X était important* :

- (24) Il était important de souligner **le rôle** de Tom Ford quand Yves Saint Laurent exerçait encore. (Corpus presse)

4.2. Le groupe 1 : le « nid » le plus propice aux Nss prototypiques

D'un point de vue général, au sein du macro-groupe (en vert dans les graphiques *supra*), les groupes 1 (en rouge) et 2 (en jaune) sont plus rapprochés l'un de l'autre, le groupe 3 (en orange) étant davantage le 3^{ème} élément de ce conglomérat. Nous allons à présent nous concentrer sur ces trois groupes, en commençant par le groupe 1, qui contient les Nss suivants : *but, mission, objectif, obstacle* et *solution*.

Le groupe 1 se démarque surtout grâce à la CS ainsi qu'à la construction à attribut de l'objet, avec une locution verbale comme *se donner pour X de + inf.* (25) ou, plus fréquemment, le verbe support *avoir* (26) :

- (25) C'est d'ailleurs ce que leur ont promis les propagandistes de l'époque et tous ceux qui **se sont donné pour mission de stimuler et de soutenir le mouvement**. (Corpus académique)

- (26) [...] les dirigeants syndicalistes (Griffuelhes, Merrheim) indiquent nettement que la multiplicité des grèves **a pour but d'imposer le problème ouvrier à la conscience bourgeoise**. (Corpus académique)

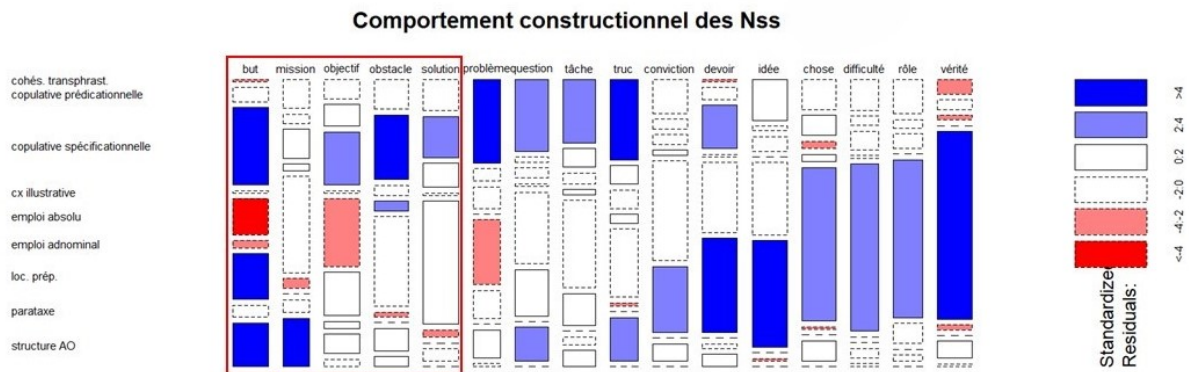


Figure 4. Les constructions privilégiées par le groupe 1 (encadrement rouge) sont les 2 constructions copulatives (spécificionnelle et prédicationnelle) ainsi que la construction à attribut de l'objet (p. ex. *avoir pour but X*).

Ce groupe semble contenir les Nss les plus « portés » vers la sous-spécification, pour reprendre les termes de Schmid : « some [shell nouns] seem to be geared for this type of usage, and can therefore be seen as prototypes of the class » ; Schmid, 2000, p. 4). En effet, les Nss de ce groupe intègrent de préférence les CS ; à l'inverse, ce groupe est négativement marqué, par rapport au macro-groupe dont il fait partie, par 2 structures : (i) les emplois adnominaux et (ii) les constructions qui établissent un lien entre le Nss et le CP via la cohésion transphrastique, comme le montre le snake plot (Figure 5) :

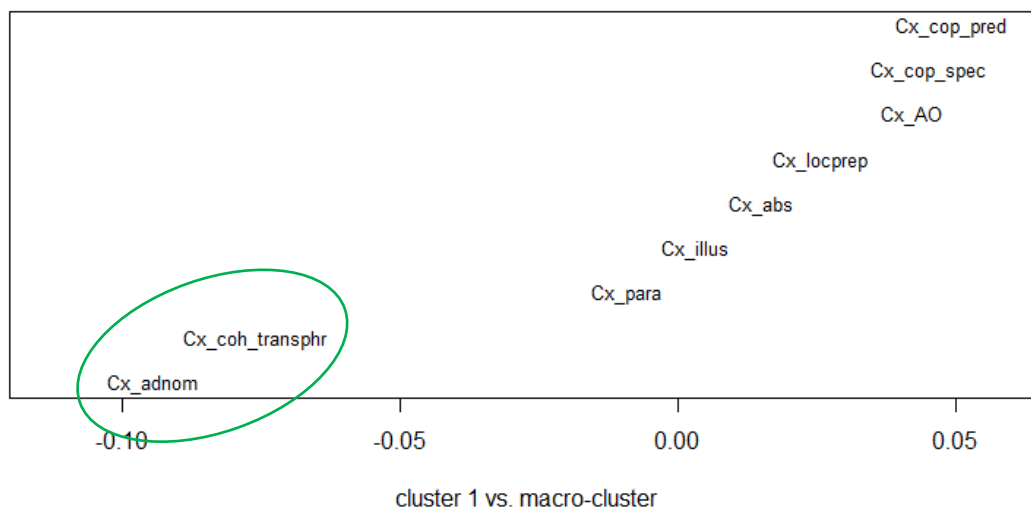


Figure 5. Les constructions les plus clivantes derrière l'opposition entre le groupe 1 et le macro-groupe sont les constructions adnominales et celles via la cohésion transphrastique.

4.2.1. Le cas particulier de *but* : un Nss en voie de figement dans une locution prépositive

Si, au sein du macro-groupe contenant les groupes 1, 2 et 3, les groupes 1 et 2 sont relativement proches, un élément semble faire « bande à part » : le Nss *but*, qui se situe entre les groupes 1 et 2. En effet, *but* se démarque par sa préférence marquée pour la construction avec locution prépositive (27) :

- (27) L'individu cherchera à se définir dans **le but** de construire son identité en tant que différente de celle de l'autre. (Corpus académique)

Le seul Nss également présent dans ce type de construction, quoique dans une proportion bien plus faible, est *objectif* (d'ailleurs un quasi-synonyme de *but*).

4.2.2. Les constructions attributives : autre marque de fabrique des Nss prototypiques

Toujours au sein du groupe 1, nous pouvons rattacher aux CS les structures à attribut de l'objet, comme l'a fait Nakamura (2012), notamment pour la locution *faire de NN* (28) :

- (28) Elle se calme, et jette un très beau sourire : « Si je vous reçois aujourd'hui, c'est parce que je veux que tout le monde sache que Dignitas existe, je veux qu'ils sachent qu'ils ont le choix de partir s'ils le souhaitent et de le faire dignement. » Ludwig Minelli en a fait **sa mission**. (Corpus presse)

Ces 2 constructions (CS « classique » et à attribut de l'objet) relèvent en effet du même mécanisme encapsulant, à savoir l'attribution d'une spécification, qui est réalisée grâce à un verbe : copule dans le cas de la CS, autre dans le cas des attributives portant sur l'objet. Grâce à ces mécanismes spécificationnels, l'« identité expérientielle », c'est-à-dire « lorsque deux éléments linguistiques ou plus contribuent à la formation d'une seule et même pensée » (Schmid, 2000, p. 29), est la plus « pure », c'est-à-dire qu'aucun doute ne subsiste quant à la relation d'encapsulation qui existe entre le Nss et son CP :

The relations of apposition and experiential identity between the noun and the postnominal clause can be tested by checking whether a matching form of the verb *be* can be inserted between them (cf. Quirk et al. 1985: 1261) without distorting the semantic relation. This is possible in example (3.1) in Figure 3.1 above: the paraphrase *the fact is that the rest of the world was against them* is compatible with the original version *the fact that the rest of the world was against them*. (Schmid, 2000, p. 30)

En termes cognitivistes, ces constructions participent à la création de catégories *ad hoc* (Legallois, 2008). La catégorisation de la portion de texte en (28) en tant que *mission* n'a donc pas vocation à durer, contrairement l'encapsulation réalisée par certaines constructions adnominales, comme en (29) :

(29) Stéphanie Converset ouvre dans sa thèse d'autres pistes de réflexion, par exemple sur le sens et les limites **du devoir d'indemnisation et de réparation du dommage par l'Etat** (lire à ce sujet le début de la deuxième partie). (Corpus académique)

La comparaison avec la construction adnominale en (29) révèle une difficulté à laquelle s'est heurté Schmid (2000), à savoir que l'identité expérientielle entre le Nss et la portion de texte suivant la préposition *de* est parfois sujette à discussion : ainsi « le devoir d'indemnisation et de réparation du dommage par l'Etat » (29) désigne un concept plus stable et durable que, par exemple, *la mission (qui est) de stimuler et de soutenir le mouvement* en (25) *supra*, qui est valable uniquement dans la situation d'énonciation dans lequel le Nss apparaît. Or, comme la pureté de l'identité expérientielle est liée à la prototypicité selon Schmid (2000), nous interprétons les Nss du groupe 1 comme formant le centre de la catégorie, qui se compose de Nss dont l'utilisation en tant que coquille encapsulante est éphémère, limitée au temps discursif.

4.3. Le groupe 2 : des Nss dédiés à une encapsulation plus « lâche »

Une fois énumérées les caractéristiques principales des Nss du groupe 1, qui, rappelons-le, forme avec le groupe 2 le couple le plus « soudé » au sein du macro-groupe dans nos données,

il convient maintenant de se pencher sur la composition du groupe 2. Ce groupe est composé des Nss suivants : *problème, question, tâche* et *truc*.

4.3.1. Les constructions qui opèrent de manière transphrastique

Le groupe 2 se démarque des autres groupes grâce à des mécanismes qui permettent une encapsulation par-delà des limites de la phrase. Ainsi, le SN *ces problèmes* en (30) fait référence à un CP dans le contexte gauche. Les Nss dont la référence opère de cette manière ont d'abord été mentionnés par Winter (1977) et Francis (1986).

- (30) Mais au terme de l'expérience exposée *supra*, il est impossible de dire si I. Ciucur :
 – « apprit » quelque chose de J. Bouët ;
 – « interpréta » une mélodie, « composée » par Brahms ;
 – « interpréta » une mélodie, « composée » par Beethoven ;
 – « interpréta » une mélodie, « composée » par lui-même (I. Ciucur) ;
 – « interpréta » une mélodie (et non, par exemple, l'« improvisa »).
- Face à **ces problèmes**, on pourrait tenter de basculer la perspective et de considérer que ce que les musiciens apprennent sont des techniques, mentales et gestuelles, leur permettant ensuite de manipuler à leur guise n'importe quelle musique (corpus académique).

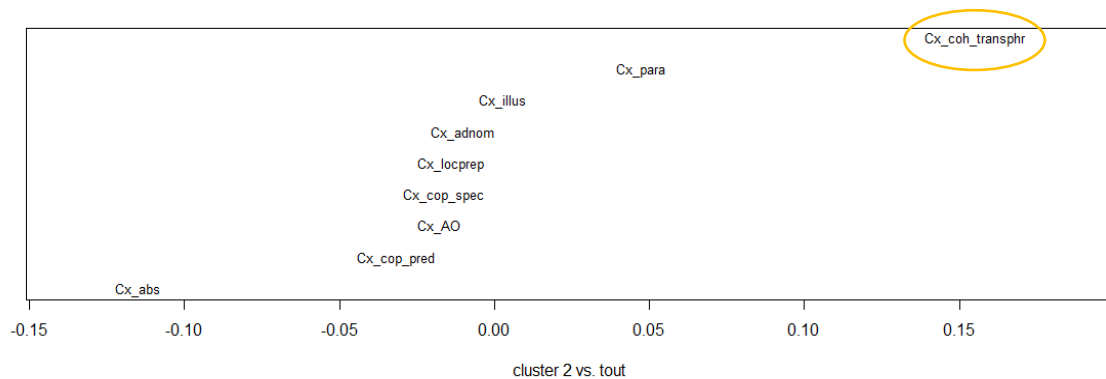


Figure 6. L'opposition entre le groupe 2 (en haut à droite) et l'ensemble des groupes est surtout due aux constructions dans lesquelles l'encapsulation se fait de manière transphrastique (entouré en jaune).

Nous considérons la construction résomptive avec un article démonstratif (CP. *Ce Nss*) comme un type de mécanisme à cohésion transphrastique. Schmid (2000) estime que, même si une relation d'encapsulation existe bien entre le Nss et son CP présent dans une autre phrase, ce lien est de nature différente que le lien qui unit les Nss à leur CP lorsque ceux-ci se trouvent dans une CS. Selon lui, affirmer que le Nss anaphorique se substitue à un CP dans le contexte

n'est pas aussi simple, car la relation qui lie le Nss au CP est moins claire que dans le cas d'une CS, où l'identité expérientielle est la plus explicite. C'est justement en voulant résoudre ce casse-tête de la référence sans un mécanisme verbal comme la copule (qui, dans une CS, constitue le « trait d'union » établissant l'identité expérientielle) que Schmid introduit le terme d'« identité expérientielle » :

Subsuming reference and anaphora as well as deixis under the idea of activation of components of a cognitive model solves the problem inherent in the claim that shell nouns and the linguistic elements expressing the shell content have to do with the same thing. On this highly general descriptive level the link between shell nouns and shell contents is that they activate identical or closely related components of a cognitive model. This co-activation is the cognitive counterpart to the pragmatic concept of co-interpretation and it is experienced by language users as experiential identity. In plain terms, experiential identity means that two or more separate linguistic elements contribute to the formation of one thought. (Schmid, 2000, p. 29)

En effet, lorsqu'un Nss fait référence à un CP dans une autre phrase, la portée de ce renvoi peut être ambiguë, par exemple parce que le CP est de nature multiple (31). C'est pourquoi, dans les cas de cohésion transphrastique, nous reprenons les termes de Cornish (1999) lorsqu'il traite de l'anaphore, en affirmant que le « déclencheur d'antécédent » (en l'occurrence, la portion de texte à interpréter comme CP) ne fait souvent que fournir « le matériau sémantique brut » (Cornish, 1999, p. 84) pour le Nss. En d'autres termes, les mécanismes qui relient un Nss à son CP via la cohésion transphrastique ne le font pas de manière aussi univoque : le rapport d'encapsulation est alors plus lâche que dans une CS, comme le montre l'exemple (31) :

- (31) a. Le temps que nous avons passé à dénicher ces parallèles, dont certains étaient prévisibles (la science-fiction) et d'autres inattendus (une théologie d'allure gnostique), est loin d'être perdu, car il nous a permis d'obtenir du danikenisme une vision plus complète que celles peintes habituellement par les pourfendeurs des parasciences. Si les commentateurs du danikenisme s'égarèrent souvent, c'est justement parce qu'ils voulaient l'expliquer avant d'en avoir saisi toute la complexité. C'est à **cette tâche préliminaire de la description** que je me suis appliqué jusqu'à présent, et il faut avouer que ce travail n'est pas encore entièrement achevé : j'ai pu établir que l'héritage de la science-fiction est présent dans toutes les versions de la théorie, mais nous ne savons pas encore s'il en va de même pour les spéculations théologiques. (Corpus académique)
- b. **Cette tâche préliminaire** est d'obtenir une vision plus complète du danikénisme / d'en saisir toute la complexité / de le décrire

4.3.2. La construction paratactique : également associée à une encapsulation plus lâche

S'il est vrai que la variante avec un démonstratif explicite davantage le lien entre le Nss et son CP que la variante avec un simple article défini, nous voyons en fait le même mécanisme à l'œuvre derrière tous les types de constructions à cohésion transphrastique, et même derrière les constructions paratactiques où la ponctuation « liante » entre Nss et CP est une ponctuation plus forte que les traditionnels « deux points », comme le point (14), ou encore le point d'interrogation (32) :

- (32) **Leur objectif ?** Obtenir le départ du président Laurent Gbagbo, qui n'assistait pas aux négociations. (Corpus presse)

Par ailleurs, au sujet d'une variante particulière de la construction paratactique, à savoir celle avec le point final, nous remarquons qu'elle s'accompagne souvent (dans 6 occurrences sur 21) d'un connecteur à vocation explicative ou par des stratégies de structuration du discours, comme en (14) (*ainsi*). Quoiqu'il en soit, la parenté entre les mécanismes à cohésion transphrastique et les constructions paratactiques est visible dans le graphique serpentin où ces deux constructions fondent l'opposition entre le groupe 2 et le groupe 3 (Figure 7). Que leur sort soit lié ressort aussi des barres dédiées à ces 2 constructions (Figure 8) :

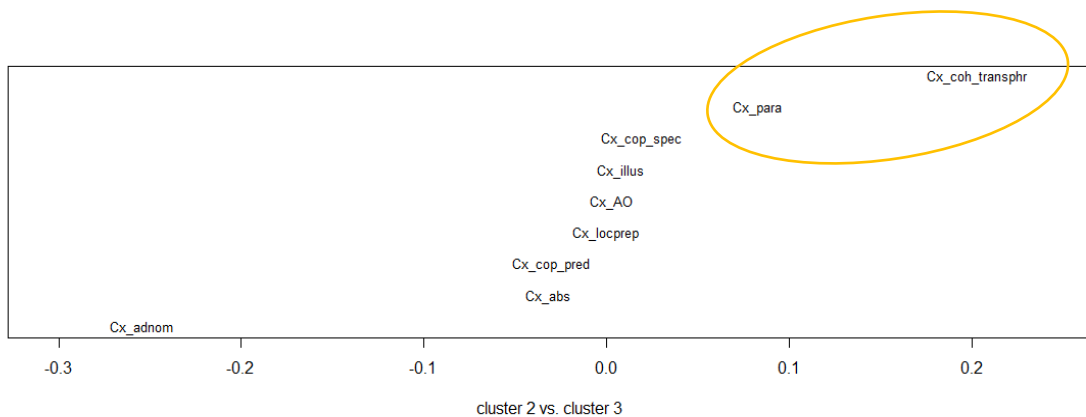


Figure 7. Snake plot montrant les constructions responsables de l'opposition entre le groupe 2 (en haut à droite) et le groupe 3 (en bas à gauche).

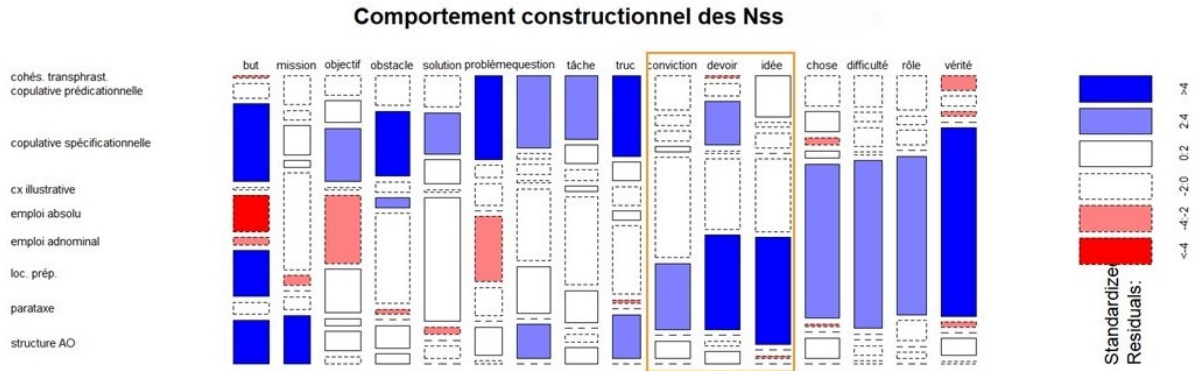


Figure 8. Les 2 constructions les plus privilégiées par les Nss du groupe 2 (encadrement jaune) sont la construction à cohésion transphrastique (tout en haut) et la construction paratactique (avant-dernière ligne).

4.4. L'opposition entre le groupe 3 et les autres groupes : une préférence globale pour l'emploi adnominal

Au sein du macro-groupe (en vert dans les figures 5, 3, 2 et 1), les groupes 1 (en rouge) et 2 (en jaune) sont plus rapprochés l'un de l'autre, le groupe 3 (en orange) venant davantage se greffer sur le couple groupe 1-groupe 2. Pour rappel, le groupe 3 contient les Nss suivants : *conviction*, *devoir* et *idée*. Il semble marqué par une préférence pour l'emploi adnominal, ce qui nous amène à un bref rappel au sujet de cette construction, plus précisément sur la manière dont l'annotation a pris en compte les différentes portées de la structure [N1 de N2].

4.4.1. Bref rappel méthodologique : la CS clivée comme test pour la portée encapsulante des structures [N1 de N2]

Dans la littérature sur les Nss, l'emploi adnominal apparaît parfois comme la variante « sans copule » de la CS : « The relations of apposition and experiential identity between the noun and the postnominal clause can be tested by checking whether a matching form of the verb be can be inserted between them [...] without distorting the semantic relation » (Schmid, 2000, p. 29). La reformulation d'une CS en une construction adnominale équivalente n'est pas toujours possible (Godard, 1996, Schmid, 2000, Legallois, 2006). Lorsqu'elle est possible, il s'agit alors selon Riegel (1996) d'une « "reclassification" contingente » ou, selon les termes de Markussen (2016) d'une structure [N1 de N2] dans laquelle le deuxième nom donne un « fond constitutif » au premier, ce qui est vérifiable grâce à la paraphrase « N2 est / constitue N1 » (Markussen 2016). C'est pourquoi nous avons utilisé la reformulation avec

une CS (clivée, de préférence) comme un test pour déterminer la portée encapsulante de la structure adnominale [N1 de N2] (Section 3.3.5).

Nous avons donc considéré les occurrences correspondant à la structure *mission + de + nominalisation*, comme en (33a), comme des emplois taxonomiques, et non comme des emplois encapsulants, en raison de l'impossible paraphrase une CS clivée (33b).

- (33) a. Toutefois, les activités de gestion administrative et de production des statistiques du chômage l'emportent largement sur **la mission de placement**. (corpus académique)
- b. ***La mission**, c'est le placement.

4.4.2. Le cas de *idée* dans la structure [N1 de N2] : des constructions adnominales bel et bien encapsulantes (Lauwers, 2012)

Le nom *idée*, tout comme *mission*, peut entrer dans la structure [N1 de N2], ce qui pourrait faire croire à un emploi taxonomique. Or, il n'en est rien, car *idée* est un « categorial term » (Lauwers, 2012) qui rentre dans une structure dans laquelle le nom nu (34a) (tout comme le nom déterminé (35a)) donne un « fond constitutif » (pour reprendre les termes de Markussen, 2016) au premier nom, comme le montrent la paraphrase avec une prédicationnelle en (34b) et les paraphrases avec des CS en (35b) et (35c). Dans les deux cas (nom nu ou déterminé), il y a une équivalence avec une structure copulative, ce qui montre que l'emploi taxonomique 'sorte de' n'est pas d'application ici, même si les deux cas de figure ne sont pas tout à fait identiques, l'identité expérientielle n'étant certainement pas aussi « pure » dans le cas d'une prédicationnelle que dans le cas d'une CS.

- (34) a. **le concept de table de décision** (exemple tiré de Lauwers, 2012)
- b. La table de décision est **un concept** déjà ancien (exemple tiré de Lauwers, 2012)
- (35) a. Une manière d'illustrer **l'idée de transmission d'un savoir**, d'un patrimoine musical au sein d'une même famille. (Corpus presse)
- b. ?**L'idée**, c'est la transmission d'un savoir.
- c. **L'idée**, c'est de transmettre un savoir.

4.4.3. Le cas de *devoir* dans la structure [N1 de N2] : des occurrences à mi-chemin entre l'emploi Nss et l'emploi taxonomique

Comme nous l'avons vu ci-dessous, il n'est pas toujours possible d'apporter une réponse claire au test de la reformulation avec une CS clivée. Ainsi, un exemple comme (29) rappelé ci-dessous passe à la fois notre test de la reformulation en une CS clivée en (29b) et le test de Markussen (2016) concernant la relation hyponymique entre N1 et N2 en (29c). Les deux

interprétations étant possibles, nous avons considéré qu'il s'agit là de cas ambigus, à mi-chemin entre les emplois réellement Nss et les emplois exprimant l'existence d'un sous-type du N tête. L'inclusion de ces cas se défend, en vertu de notre politique d'inclusion maximale (Section 3.3.5), même s'il est vrai qu'ils se trouvent en marge du spectre :

- (29) b. **Le devoir de l'État**, c'est d'indemniser et de réparer le dommage
 c. Le devoir d'indemnisation et de réparation du dommage par l'État est un type de devoir

4.4.4. Bilan intermédiaire pour les Nss du groupe 3 : un emploi adnominal parfois ambigu

Comme la CS est l'une des principales caractéristiques du groupe 1, nous allons non seulement comparer à quel point la construction adnominale justifie l'opposition entre le groupe 3 et le reste des groupes (Figure 9), mais également à quel point elle justifie l'opposition entre le groupe 3 et le groupe 1 (Figure 11).

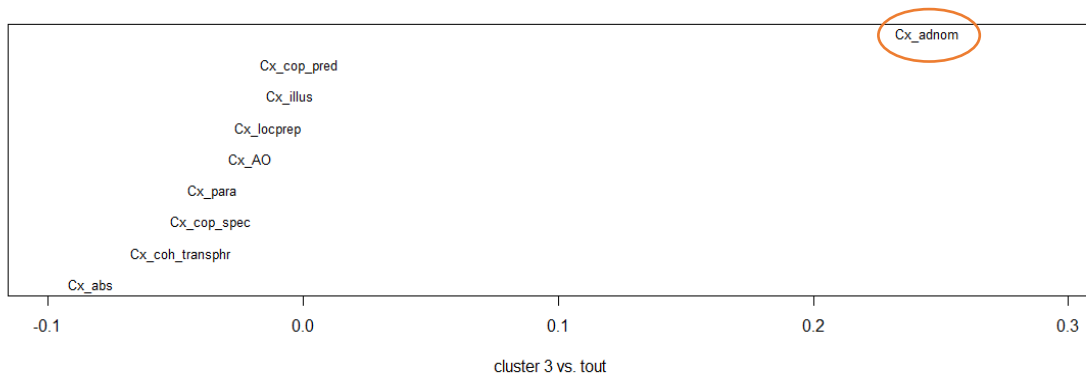


Figure 9. L'emploi adnominal est la construction la plus clivante derrière l'opposition entre les Nss du groupe 3 (en haut à droite) et l'ensemble des groupes (en bas à gauche).

En effet, non seulement l'emploi adnominal permet d'opposer le groupe 3 aux autres groupes (Figure 9), il représente aussi le contexte le plus fui par les Nss du groupe 1 (les barres rouges dans la Figure 10) :

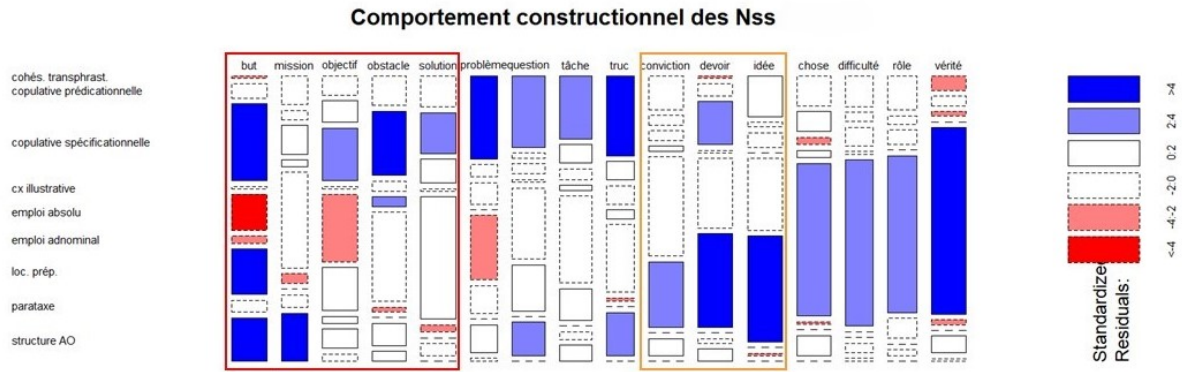


Figure 10. S'il y a une construction que les Nss du groupe 3 (encadrement orange) préfèrent et que les Nss du groupe 1 (encadrement rouge) évitent, c'est l'emploi adnominal.

L'emploi adnominal est également la construction qui participe majoritairement à l'opposition entre le groupe 3 et le groupe 1, comme le montre le graphique serpentin de la Figure 11 :

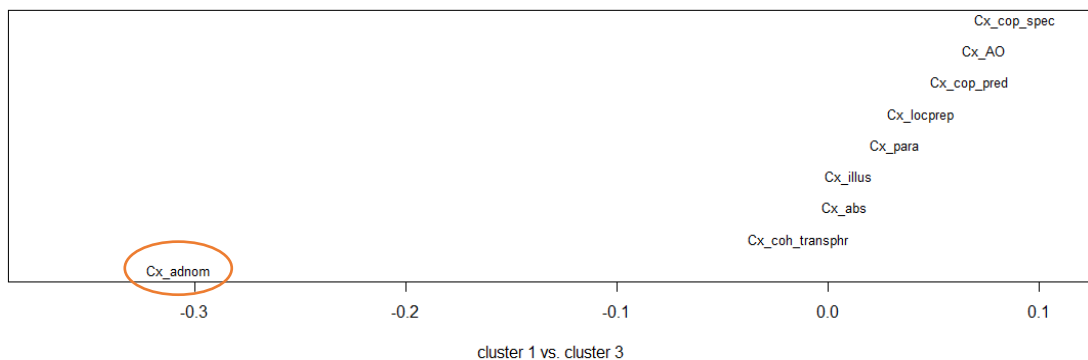


Figure 11. L'emploi adnominal participe davantage à l'opposition entre le groupe 3 (en bas à gauche) et le groupe 1 que ne le fait la construction attributive.

Alors que la CS concerne aussi bien les Nss du groupe 1 que du groupe 3 (quoique dans une mesure nettement plus faible), la construction adnominale se spécialise dans les Nss du groupe 3, dont l'« identité expérientielle » n'est, le plus souvent, pas aussi « pure » que celle des Nss dans une CS. La construction adnominale, de par la jonction du CP (potentiel) au Nss (potentiel), donne une fonction classifiante au complément, créant ainsi un sous-type qui se rapproche de l'emploi taxonomique (36c) :

- (36) a. D'une part, nous devons faire face à **une difficulté de maintenance**, liée à l'obsolescence du langage Cobol. (Corpus discours & allocutions)
- b. #La difficulté, c'est de maintenir
- c. La difficulté de maintenance est un type de difficulté (test fourni par Markussen, 2016)

Cette fonction parfois classifiante est justement ce qui permet de départager les quasi-synonymes *mission* et *devoir*, *mission* préférant surtout les constructions typiques du groupe 1, alors que *devoir* se spécialise dans cet emploi quasi-taxonomique :

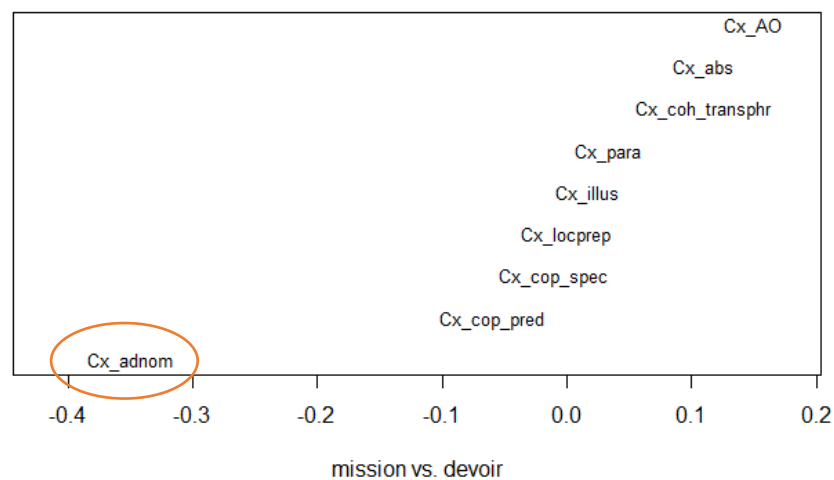


Figure 12. Les spécialisations constructionnelles des quasi-synonymes *mission* (en orange) et *devoir* révèlent une portée pragmatique différente, le sémantisme des Nss étant par définition variable et donc moins « inévitable » que celui des noms composés.

En termes de répartition dans les 6 genres textuels qui composent notre corpus, on remarque que *devoir* dans la construction adnominale est surtout présent dans les sous-corpus « académique » et « discours & allocutions », dans lesquels les locuteurs mobilisent des stratégies discursives différentes que dans les corpus plus spontanés. En formant un nom composé lui-même (possiblement) porteur de l'action qui doit être réalisée (37), le locuteur donne des instructions (*devoir* étant un Nss déontique) de manière plus détournée que s'il avait clairement assigné un CP au Nss, via une CS par exemple. La condensation d'un CP en un SN composé pourrait également venir de la préférence des genres académiques en particulier pour la nominalisation, qui permet un « empaquetage » de l'information (Halliday & Webster, 2004).

- (37) En revanche, lorsque la victime a fait usage du droit de refuser de témoigner sur des faits qui concernent sa sphère intime (art. 84e III CPP/36 II LAVI) dans le cadre

de la procédure pénale, l'autorité ne peut lui reprocher une violation de **son devoir de collaborer**. (Corpus académique)

5. Résultats de l'analyse en composantes principales : les constructions qui structurent le champ des Nss

Dans ce qui précède, la discussion des différents groupes nous a déjà amenés à identifier les constructions constitutives de tel ou tel groupe. À cet effet, nous avons analysé les graphiques serpentins, ainsi que l'analyse des résiduels issus de l'analyse bivariée croisant les Nss avec les constructions majeures.

Pour analyser de manière plus globale les relations entre ces constructions dans la constitution de la classe des (emplois) Nss, nous avons réalisé une analyse en composantes principales. Notons que les données pour cette analyse ont été normalisées, afin de contrer des effets de taille :

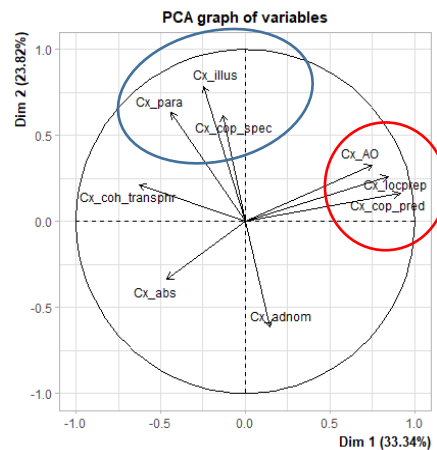


Figure 13. L'analyse en composantes principales (ACP) a identifié 2 faisceaux de constructions au comportement analogue.

Dans l'ensemble, chaque construction dans le graphique *supra* semble présenter un comportement *sui generis*, même s'il est possible de regrouper certaines d'entre elles sous deux faisceaux de constructions majeurs. Parmi les constructions du premier faisceau (en bleu), nous n'avons pas encore mentionné la construction illustrative (38), en raison de sa faible fréquence d'apparition. Cette construction permet au Nss d'encapsuler cataphoriquement une portion de texte, propriété partagée avec les CS, ce qui pourrait expliquer la proximité des deux flèches sur le graphique *supra*.

- (38) En suivant l'argumentaire préparé par la CGT, le défenseur de E. déconstruit l'ensemble des équivalences présumées par la sanction, en signalant tous **les obstacles**, toutes les causes de ralentissement qui invalident la mesure de ses

performances. Par exemple, si la présence du siège soulage E. de ses douleurs (la position debout prolongée étant contre-indiquée), il va beaucoup moins vite que les salariés qui n'ont pas de siège. (Corpus académique)

Ce faisceau (en bleu) semble marqué par une structure bipartite, avec un premier segment, composé du Nss, qui engendre une attente, et un deuxième segment, composé du CP, qui vient répondre à cette attente. Cette structure est privilégiée par les CS et par certaines constructions paratactiques. Ce premier faisceau réunit donc les constructions dont le fonctionnement est le plus assimilable à celui des pseudo-clivées (39), rapprochement soutenu par Legallois et Gréa (2006) et par Apothéloz et Roubaud (2015), ces derniers estimant même que les CS et les constructions pseudo-clivées relèvent du même mécanisme.

(39) **Ce qui me plaît** c'est la linguistique (exemple tiré de Roubaud, 2000)

Le deuxième faisceau (en rouge) réunit les constructions prédicationnelles, les constructions à verbe support (26) et les locutions prépositives (27). Les constructions à cohésion transphrastique, les emplois absolus et nominaux font « bande à part ». Ils semblent avoir un comportement *sui generis*, qui est difficile à rattacher aux autres constructions.

6. Conclusion

Nos analyses outillées suggèrent l'existence de 4 groupes de Nss parmi les 16 Nss de notre corpus. Cette catégorisation nous permet ainsi de répondre à la première question de recherche énoncée en introduction, à savoir :

1) Quels Nss sont concernés par un comportement distributionnel similaire en termes d'intégration dans ces constructions ?

Le premier groupe est composé de Nss qui s'intègrent de manière privilégiée dans des constructions établissant un lien explicite entre le Nss et son CP, à savoir les constructions copulatives (spécificationnelles et prédicationnelles) et les constructions à attribut de l'objet (*avoir pour but de + inf*). Avec Schmid, on peut dire que ces Nss correspondent au fonctionnement prototypique du Nss. Le deuxième groupe, assez proche du premier, regroupe ensuite des Nss qui intègrent des constructions à encapsulation plus « lâche », comme les constructions résomptives (*ce Nss*) ou les constructions paratactiques bipartites (*le Nss ...*).

Ensuite, plus on s'éloigne de ces deux groupes, plus le lien entre le Nss et son CP s'éloigne de la relation expérientielle pure nécessaire pour parler de véritable encapsulation. On retrouve ainsi, dans le troisième groupe, des Nss comme *conviction*, *devoir* et *idée*, qui s'intègrent plus souvent que les autres dans des constructions adnominales, susceptibles de participer à la création de noms composés dotés d'une forte cohésion interne. Ce groupe s'oppose, en termes de préférence pour la construction adnominale, au groupe 1. Enfin, en

périphérie (groupe 4), nous retrouvons des emplois dits « absolus », auto-saturés, qui se trouvent à la limite du spectre Nss. C'est par exemple le cas des noms abstraits en emploi générique (*la vérité*) ou des quasi-pronoms (*choses*).

Notre deuxième question de recherche était la suivante :

2) Ces catégories distributionnelles correspondent-elles à des types sémantico-conceptuels de Nss ?

Il nous paraît impossible de voir dans les groupes ainsi formés un regroupement sur la base d'une éventuelle portée modale des Nss (Adler & Legallois, 2018, et Legallois & Vajnovszki, 2021), même si la plupart des Nss du groupe 1 sont des bouliques (excepté *obstacle*, un appréciatif négatif). Nous pouvons néanmoins mieux cerner la manière dont certains Nss au sens proche fonctionnent : c'est ainsi que *mission* fait partie du groupe 1 (celui des Nss privilégiant les constructions créant une identité expérientielle « pure » entre le Nss et son CP), alors que *devoir* fait partie du groupe 3 (celui se démarquant par le nombre d'emplois quasi-taxonomiques, p. ex. *le devoir d'indemnisation*) ; de même, *problème* fait partie du groupe 2 (groupe marqué pour sa préférence pour les constructions à encapsulation plus lâche), alors que *obstacle* préfère les constructions où l'encapsulation est univoque (groupe 1).

C'est pourquoi ces résultats nous permettent également de répondre par la négative à la quatrième question de recherche, qui consistait à se demander :

4) Une synonymie entraîne-t-elle une appartenance à la même catégorie ?

Au contraire, comme mentionné plus haut, certains quasi-synonymes semblent se démarquer entre eux par une intégration privilégiée dans l'une ou l'autre construction. À ce sujet, une étude diachronique sur le comportement de ces quasi-synonymes pourrait permettre de vérifier si le caractère parfois concret du Nss *obstacle* (par rapport à *problème*, qui signifie littéralement 'ce que l'on jette devant soi') est précisément ce qui gêne encore son intégration dans des constructions comme la résomptive (*cet obstacle*).

Concernant la troisième question de recherche, celle-ci était :

3) Cette catégorisation correspond-elle à l'existence d'un prototype ?

Comme nous l'avons déjà indiqué en réponse à la question 1, notre analyse nous a permis de mettre au jour certains Nss qui correspondent maximale-ment au fonctionnement prototypique du Nss, à savoir ceux qui entrent dans des mécanismes qui établissent une identité expérientielle plus « pure » entre le Nss et son CP (selon les termes de Schmid) : les CS et les constructions attributives. Puis, de proche en proche, nous avons identifié des groupes de Nss de moins en moins prototypiques, nous conduisant jusqu'aux marges du spectre. Rappelons toutefois que, comme précisé en introduction, notre analyse n'avait pas pour but

de dévoiler les propriétés prototypiques du Nss : elle a simplement mené à l'identification d'éléments correspondant maximale à cette définition prototypique.

Enfin, nous reproduisons la cinquième et dernière question de recherche ci-dessous :

5) Existe-t-il des liens entre les constructions d'accueil ?

Une analyse en composantes principales nous a permis d'identifier 2 faisceaux de constructions au fonctionnement similaire, à savoir les constructions bipartites, proches des pseudo-clivées (comme les CS, les paratactiques avec les deux points encapsulants, etc.) et les autres constructions (construction à attribut de l'objet, locution prépositive *etc.*).

Le comportement des Nss par rapport aux constructions majeures, à un niveau de granularité assez grossier, n'épuise bien entendu pas la problématique. Dès lors, les résultats de la présente étude demandent à être complétés par des analyses portant sur d'autres facettes du comportement Nss, par exemple sur le type de CP (infinitif, SN, etc.) ou sur leur affinité par rapport à la modification par des épithètes.

Bibliographie

- Adler, S. (2017). Les noms généraux – « shell nouns » – participent-ils à une lecture taxinomique de type Hiérarchie-être ? *Syntaxe et sémantique*, 1(18), 45-66.
- Adler, S., & Legallois, D. (2018). Les noms sous-spécifiés dans le débat parlementaire : Analyse fréquentielle et catégorisation modale. *Langue française*, 198(2), 19-34. <https://doi.org/10.3917/lf.198.0019>
- Apothéloz, D., & Roubaud, M.-N. (2015). Constructions pseudo-clivées. In *Encyclopédie grammaticale du français*. <http://encyclogram.fr>
- Benninger, C. (2014). La question de la définition sémantique du nom atypique chose. *Travaux de linguistique*, 69, 35-55. <https://doi.org/10.3917/tl.069.0035>
- Blanche-Benveniste, C. (1992). Sur un type de nom évaluatif portant sur des séquences verbales. *Review of applied linguistics*, 97(1), 1-25. <https://doi.org/10.1075/ital.97-98.01bla>
- Cornish, F. (1999). *Anaphora, Discourse, and Understanding: Evidence from English and French*. Clarendon Press.
- Flowerdew, J. (2002). A pedagogic grammar of signalling nouns in discourse. *Revista canaria de estudios ingleses*, 44, 141-55.
- Flowerdew, J., & Forest, R. W. (2014). *Signalling nouns in English*. Cambridge University Press.
- Francis, G. (1986). *Anaphoric Nouns. Discourse Analysis Monographs 11*. University of Birmingham.

- Francis, G. (1994). Labelling discourse: an aspect of nominal-group lexical cohesion. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 83-101). Routledge.
- Godard, D. (1996), « Les phrases compléments de Nom sont-elles des arguments ? ». In N. Flaux, M. Glatigny & D. Samain (Eds.), *Les noms abstraits. Histoire et théories* (pp. 301-312). Presses Universitaires du Septentrion.
- Gries, S. T. (2008). Behavioral Profiles 1.0. A program for R 2.7.1 and higher.
- Halliday, M. A. K., & Webster, J. (2004). *The language of science*. Continuum.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Routledge.
- Ho-Dac, L. M., Miletic A., Wauquier M., & Fabre, C. (2020). Approches outillées pour l'étude des noms sous-spécifiés ou noms capsules. *Le Français Moderne*, 1, 44-63.
- Huyghe, R. (2015). Les typologies nominales : présentation. *Langue française*, 185, 5-27. <https://doi.org/10.3917/lf.185.0005>
- Ivanič, R. (1991). Nouns in search of a context: a study of nouns with both open- and closed-system characteristics. *International Review of Applied Linguistics*, 29(2), 93-114. <https://doi.org/10.1515/iral.1991.29.2.93>
- Kassambara, A. (2017). *Practical guide to principal component methods in R: PCA, M(CA), FAMD, MFA, HCPC, factoextra*. STHDA (Statistical tools for high-throughput data analysis). <http://www.sthda.com/english/wiki/practical-guide-to-principal-component-methods-in-r>
- Kassambara, A., & Fabian, M. (2020). *Factoextra: Extract and visualize the results of Multivariate Data Analyses*. R package version 1.0.7. <https://cran.r-project.org/web/packages/factoextra/readme/README.html>
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: an introduction to cluster analysis*. Wiley.
- Kleiber, G. (1987). Mais à quoi sert donc le mot *chose* ? Une situation paradoxale. *Langue française*, 73, 109-128. <https://doi.org/10.3406/lfr.1987.6431>
- Langacker, R. (1987). *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford University Press.
- Lauwers, P. (2021). Plurality Without (Full) Countability: On Mass-Like Categories in Lexical Plurals. In T. Kiss, F.-J. Pelletier & H. Husić (Eds.), *Things and Stuff: The Semantics of the Count-Mass Distinction* (pp. 337-356). Cambridge University Press.
- Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* 25(1), 1-18.

- Lauwers, P. (2012). Does French really have bare singular subject NPs? On semantic reconstruction partly based on (cata)phoric ellipsis. *Neuphilologische Mitteilungen*, 113, 3-40.
- Lefeuvre, F. (2007). Le segment averbal comme unité textuelle. In M. Charolles (Ed.), *Parcours de la phrase : mélanges offerts à Pierre Le Goffic* (pp. 143-158). Ophrys.
- Legallois, D., & Gréa, P. (2006). L'objectif de cet article est de... Construction spécifique et grammaire phraséologique. *Cahiers de praxématique*, 46, 161-186. <https://doi.org/10.4000/praxematique.657>
- Legallois, D. (2008). Sur quelques caractéristiques des noms sous-spécifiés. *Scolia*, 23, 109-127.
- Legallois, D., & Vajnovszki, A. (2021). Proposition pour une conception modale des noms dits sous-spécifiés. *CORELA*, HS-34. <https://doi.org/10.4000/corela.13617>
- Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. John Benjamins.
- Markussen, M. (2016). *La construction <Nom + de + Nom> en français contemporain : essai d'une approche cognitive*. Presses Universitaires de Caen.
- Meyer, D., Zeileis A., & Hornik, K. (2023). *vcd: Visualizing Categorical Data*. R package version 1.4-11. <https://cran.r-project.org/package=vcd>
- Nakamura, T. (2012). Une construction à double complément du verbe « faire » : attribut de l'objet indirect ? *3^{ème} Congrès Mondial de Linguistique Française*, Jul 2012, Lyon, France, 2393-2414. <https://doi.org/10.1051/shsconf/20120100249>
- Riegel, M. (1996). Les noms à compléments propositionnels : en quoi sont-ils plus abstraits que les autres. In N. Flaux, M. Glatigny & D. Samain (Eds.), *Les noms abstraits. Histoires et théories* (pp. 313-322). Presses Universitaires du Septentrion.
- Roubaud, M.-N. (2000). *Les constructions pseudo-clivées en français contemporain*. Honoré Champion.
- Roze, C., Charnois, T., Legallois, D., Ferrari, S., & Salles, M. (2014). Identification des noms sous-spécifiés, signaux de l'organisation discursive. *Proceedings of TALN 2014*, 1, 377-388.
- Schmid, H.-J. (2000). *English Abstract Nouns As Conceptual Shells: From Corpus to Cognition*. Mouton De Gruyter.
- Tadros, A. (1994). Predictive categories in expository text. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 68-82). Routledge.

Van Wettere, N. (2018). *Copularité et productivité : une analyse contrastive des verbes attributifs issus de verbes de mouvement en français et en néerlandais*. Thèse de doctorat, Université de Gand.

Vendler, Z. (1968). *Adjectives and nominalizations*. Mouton De Gruyter.

Winter, E. O. (1977). A clause-relational approach to English texts: A study of some predictive lexical items in written discourse. *Instructional Science*, 6(1), 1-92.
<https://doi.org/10.1007/BF00125597>

Winter, E. O. (1982). *Towards a contextual grammar of English: the clause and its place in the definition of sentence*. Routledge.

Winter, E. O. (1992). The notion of unspecific versus specific as one way of analysing the information of a fund-raising letter. In W. C. Mann & S. A. Thompson (Eds.), *Discourse Descriptions. Diverses Analyses of a Fund-Raising Text* (pp. 131-170). John Benjamins.