

Application d'une méthodologie d'analyse des prédicats nominaux : l'exemple du lexème MORT₁

Carlos VALCÁRCEL RIVEIRO

Universidade de Vigo

carlos.valcarcel@uvigo.es

<https://orcid.org/0000--0003-1123-5211>

Laura PINO SERRANO

Universidade de Santiago de Compostela

laura.pino@usc.es

<https://orcid.org/0000--0003-1276-0966>

Resumen

Nuestro trabajo plantea una metodología para analizar predicados nominales, centrándose en el predicado *muerte* en el sentido de 'fin de la vida'. Destacamos la complejidad de los grupos nominales y su polisemia y proponemos un enfoque dependencial para comprender mejor su funcionamiento interno. Los datos se recogieron a partir de un corpus y requirieron un tratamiento selectivo antes de ser anotados sintáctico-semánticamente. Este análisis nos permitió identificar varios aspectos relevantes, como la centralidad de los argumentos nominales. Dada la utilidad de este enfoque para la enseñanza de idiomas y la lexicografía, proponemos por último continuar esta investigación para otros tipos de predicados y experimentar con otros sistemas de anotación semántica.

Palabras clave: predicado nominal, gramática dependencial, complementación, combinatoria, metodología de análisis.

Résumé

Notre travail consiste à présenter une méthodologie pour analyser les prédicats nominaux en nous concentrant sur le prédicat *mort* au sens de 'cessation de la vie'. Nous soulignons la complexité des groupes nominaux et de leur polysémie et proposons une approche dépendancielle pour mieux comprendre leur fonctionnement interne. Les données ont été recueillies à partir d'un corpus et ont nécessité un dépouillement sélectif avant d'être annotées syntaxico-semantiquement. Notre analyse a permis de relever plusieurs questions pertinentes, dont la centralité des arguments au sein d'un prédicat nominal. Étant donné l'utilité de cette démarche pour l'enseignement des langues et la lexicographie, nous proposons finalement de poursuivre ces recherches sur d'autres types de prédicats et de tester d'autres systèmes d'annotation sémantique.

Mots clé : prédicat nominal, grammaire dépendancielle, complémentation, combinatoire, méthodologie d'analyse.

* Artículo recibido el 27/04/2023, aceptado el 29/06/2023.

Abstract

Our paper proposes a methodology for analysing nominal predicates, focusing on the predicate *death* meaning ‘end of life’. We highlight the complexity of nominal groups and their polysemy, and propose a dependency approach to better understand their inner workings. Data were collected from a corpus and required selective treatment before being syntactically-semanticly annotated. This analysis allowed us to identify several relevant aspects, such as the centrality of nominal arguments. Given the usefulness of this approach for language teaching and lexicography, we propose to continue this research for other types of predicates and to experiment with other semantic annotation systems.

Keywords: nominal predicate, dependency grammar, complementation, combinatorics, analysis methodology.

0. Justification de la recherche

Quelle est la complexité des mots qui nous entourent ? Nous avons tous déjà été confrontés à des noms qui ont plusieurs significations, comme MORT par exemple. Mais quelles sont les implications de cette polysémie sur notre compréhension des mots et comment pouvons-nous les étudier ? C’est ce que nous avons cherché à découvrir dans notre recherche sur le prédicat nominal MORT, c’est-à-dire l’ensemble des compléments essentiels qui sont nécessaires pour en déterminer le sens.

Maints travaux ont été publiés ces dernières décennies à propos des problématiques liées à la complémentation nominale (Condette *et al.*, 2012 ; Gross, 1991, 2012 ; Stage, 1997). La plupart de ces études soulignent la complexité des relations des actants autour du groupe du nom (GN) et visent surtout des questions d’ordre syntaxique ou formel : position et combinatoire des constituants du groupe nominal, restrictions dans l’emploi des déterminants et des prépositions, transpositions syntaxiques des nominalisations déverbales, etc. Pourtant, une approche plus sémantique de ces aspects est encore nécessaire pour essayer d’apporter un peu plus de lumière sur les principes et les contraintes qui déterminent le fonctionnement interne du GN. De ce point de vue, nous nous interrogeons à propos de la pertinence d’analyser en détail le sémantisme des arguments nominaux pour essayer de comprendre toutes ces questions.

Pour ce faire, nous avons choisi le prédicat nominal MORT pour entreprendre notre recherche syntaxico-sémantique. Ce substantif fait partie de la liste des vingt noms étudiés dans le cadre des projets de recherche MultiGenera¹ et MultiComb², où nous avons développé la méthodologie que nous appliquons dans ce travail. Nous avons

¹ Le projet MultiGenera a été subventionné par la Fondation BBVA dans le cadre de son programme d’aides aux équipes de recherche scientifique de 2017 (Domínguez, Valcárcel et Lindemann, 2018).

² Le projet MultiComb (FFI2017-82454-P) a été financé par les fonds FEDER et par le Ministerio de Economía, Industria y Competitividad — Agencia Estatal de investigación (Domínguez, 2022).

sélectionné un prédicat biargumental pour notre étude étant donné le détail des analyses à réaliser pour chaque argument. Ainsi, il s'agira non seulement de déterminer quels sont ces deux arguments, mais aussi de définir leur rôle sémantique, leurs différentes réalisations formelles et les étiquettes sémantiques y associées.

Notre travail se divise en plusieurs parties : dans un premier temps, nous définirons les sens retenus pour notre analyse, puis nous décrirons notre méthodologie d'analyse des groupes nominaux. Ensuite, nous étudierons les données recueillies sur le prédicat MORT, en mettant en évidence les arguments expérient et cause, ainsi que les réalisations formelles associées. Enfin, nous conclurons notre travail en discutant des résultats obtenus et des perspectives de recherche futures.

1. Cadre théorique

1.1. La complexité du groupe nominal

Les relations des constituants au sein du groupe nominal (GN) sont bien plus complexes que celles existantes dans le groupe verbal (GV) ou la phrase où les éléments ont une place plus ou moins fixe qui détermine, dans bien des cas, leur rôle et leur fonction par rapport au prédicat verbal (Condette *et al.*, 2012 ; Gross, 2012 : 102-103, 106 ; Stage, 1994).

Par contre, dans le GN les constituants se situent plus librement à droite ou à gauche du noyau et parfois il n'est pas facile de déterminer leur rôle, leur fonction ou leur combinatoire. D'ailleurs, tout comme dans le cas du verbe (Mel'čuk, 2015, 5), il n'existe pas non plus de consensus entre les spécialistes lorsqu'il s'agit de distinguer ce qui est actantiel ou non au sein du GN. Ainsi, pour certains auteurs comme Mel'čuk (2015 : 120), les adjectifs (par exemple « la présence française ») ne correspondent pas à des réalisations actantielles, mais circonstancielles. Le comportement syntaxique de surface des adjectifs est déterminant pour ces spécialistes : les adjectifs apparaissent dans des constructions directes et ne sont pas précédés d'une préposition marquant une réaction nominale. Cependant, pour d'autres auteurs, des critères purement formels ne suffisent pas à définir les actants et ils considèrent les déterminants possessifs et les adjectifs relationnels comme des réalisations actantielles sur la base de critères sémantiques (Domínguez, 2011 : 140-146 ; Rigau, 1999 : 339-352). Signalons ici que Mel'čuk (2015 : 120) reconnaît lui-même que des réalisations non actantielles au niveau formel ou « de surface » constituent des représentations d'actants sémantiques « profonds » et donc des participants à des prédicats nominaux.

En effet, le GN est beaucoup plus complexe qu'il ne le paraît à première vue, étant donné le jeu des prépositions pertinentes ainsi que le choix des déterminants possibles ou non dans chacun des cas (Gross, 1991 : 269-270). Déceler les raisons pour lesquelles il existe ce type de restrictions, bien plus fréquentes qu'au sein du GV, est une tâche aussi passionnante que compliquée. Pour ce faire, il faut indiquer que la

sélection sémantique des différents arguments faisant partie du GN est d'une importance majeure et qu'elle joue un rôle prioritaire dans la complémentation nominale et dans sa combinatoire.

Par conséquent, l'étude des prédicats nominaux, tout comme les verbaux, comporte l'utilisation de catégories d'analyse sémantiques. Tout d'abord, chaque argument a un certain rôle sémantique, qui le définit par rapport au noyau du groupe nominal dans lequel il se trouve et qui, en même temps, détermine la sélection lexicale de son propre noyau. Ainsi, les candidats lexicaux susceptibles d'occuper un noyau argumental doivent présenter des caractéristiques sémantiques en adéquation avec le rôle qu'ils joueront dans le groupe du nom.

Par exemple, le lexème RÉFLEXION ('capacité de la pensée') ne peut pas former le noyau de l'argument agentif du nom VOYAGE ('déplacement d'un être animé sur une certaine distance'), puisqu'il ne possède pas le trait +animé dans son sémantisme, qui est requis pour y jouer un rôle agentif. De ce fait, et même s'il s'agit d'une structure grammaticale d'un point de vue strictement formel, il est sémantiquement inacceptable de dire « *le voyage de la réflexion en Italie » (*la réflexion voyage en Italie'). Pourtant, il serait tout à fait possible de dire « le voyage de réflexion en Italie » ('le voyage consacré à la réflexion'), puisque, sans le déterminant, nous avons déjà affaire à un autre type d'argument à rôle sémantique de classe et pour lequel le sémantisme de RÉFLEXION est adéquat.

En fait, lorsque deux arguments du même prédicat présentent des réalisations formelles identiques, c'est précisément le sémantisme du noyau argumental qui nous permet d'identifier le rôle attribué dans de nombreux cas³. Par exemple, dans la phrase « la discussion de la directrice avec M. Robert », nous savons que DIRECTRICE joue un rôle agentif, tandis que dans « la discussion du budget avec M. Robert », nous interprétons que BUDGET a un rôle sémantique de 'thème'.

Évidemment, les traits sémantiques associés à un certain rôle argumental changent d'un prédicat à l'autre. Ainsi, le rôle agentif dans le prédicat ODEUR ('émanation perçue par le nez') permet un plus large éventail de traits sémantiques que dans le prédicat VOYAGE. Par exemple, les lieux jouent un rôle agentif dans le cas de ODEUR (l'odeur de la chambre), mais non pas dans celui de VOYAGE (*le voyage de la chambre).

La théorie des classes d'objets née au sein du Laboratoire de Linguistique Informatique (LLI) sous la direction de G. Gross et ses disciples, ainsi que d'autres approches théoriques semblables ou complémentaires, peuvent servir à l'identification et à l'explication de la formation des GN et à leur caractérisation. Ce courant théorique nous guidera dans l'étude du prédicat nominal MORT lorsque nous essaierons d'analyser les types d'arguments possibles ou interdits dans chacune de ses réalisations.

³ Les interprétations ambiguës sont aussi fréquentes. Par exemple, dans la phrase « la vidéo des enfants », ENFANTS peut être interprété comme le noyau d'un argument soit à rôle agentif ('les enfants ont fait la vidéo'), soit à rôle 'thème' ('la vidéo porte sur les enfants').

1.2. L'annotation sémantique : ontologies, classification et étiquetage.

Comme dans le cas des prédicats verbaux, étant donné la relation étroite entre les réalisations formelles des compléments nominaux et leur rôle sémantique, il est essentiel de prendre en compte les caractéristiques sémantiques au sein de chaque argument. Pour ce faire, le recours à un outil d'analyse sémantique s'avère incontournable. Dans le cadre des deux projets dans lesquels nous avons entrepris notre analyse, on a envisagé plusieurs possibilités : des ontologies, des classifications et des systèmes d'étiquetage sémantiques.

La première option considérée visait les ontologies sémantiques. En fait, ces outils d'analyse constituaient déjà une partie essentielle du travail dans les projets MultiGenera et MultiComb. On les utilisait pour extraire de façon semi-automatique des sélections lexicales afin de couvrir l'axe paradigmatique des arguments nominaux à gérer (Domínguez, Valcárcel et Lindemann, 2018). À cette fin, on utilisait le modèle EuroWordNet du Multilingual Central Repository (MCR)⁴, où les données lexicales sont associées à des caractéristiques sémantiques ou cognitives catégorisées dans différentes ontologies. En particulier nous avons travaillé avec Suggested Upper Merged Ontology⁵ (SUMO), Top Concept Ontology⁶ (Top), WordNet Domains⁷, Basic Level Concept⁸ et Epinonyms (Domínguez, Solla et Valcárcel, 2019; Gómez et Solla, 2018). Cependant, on les a évitées pour l'étiquetage sémantique. S'il est vrai que ces ontologies ont été extrêmement utiles pour l'extraction lexicale massive, elles ne se prêtent pas aussi bien à l'étiquetage des paquets lexicaux résultants. Ces paquets sont à la base de la génération automatique d'arguments nominaux et, par conséquent, leur élaboration et leur étiquetage sont soumis à des contraintes combinatoires que nous illustrerons pour le cas de MORT. Pourtant, les catégories des ontologies utilisées répondent plus à des critères cognitifs que sémantiques⁹.

La deuxième option considérée visait sur les classifications sémantiques. Dans le cadre des projets qui nous occupent, on est parti d'une classification générale appliquée par M^a José Domínguez (2011) et adoptée pour le développement du dictionnaire

⁴ <http://adimen.si.ehu.es/web/MCR>

⁵ <http://www.adampease.org/OP/>

⁶ http://globalwordnet.org/gwa/ewn_ewn_to_bc/ewnTopOntology.htm

⁷ <http://wndomains.fbk.eu/>

⁸ <http://sli.uvigo.gal/galnet/index.php?lg=en>

⁹ Un exemple clair de cette inadéquation des catégories ontologiques pour l'étiquetage des paquets lexicaux est celui des noms de pays. Dans les catégories ontologiques proposées dans le MCR ce champ sémantique est regroupé normalement sous une même catégorie. Cependant, étant donné les contraintes formelles liées au genre des noms des pays en français, on a dû créer plusieurs paquets lexicaux et donc des étiquettes ad hoc. Par contre, l'étiquette du paquet lexical *animado_humano_cargo* regroupe des lexèmes tirés de plusieurs catégories ontologiques. Pour le cas de l'ontologie SUMO les catégories associées à ce paquet sont *SocialRole*, *OccupationalRole* et *Position* parmi d'autres.

PORTLEX (Domínguez et Valcárcel, 2020). Il s'agit, pourtant, d'une classification traditionnelle similaire à celle que l'on peut retrouver dans les grammaires françaises.

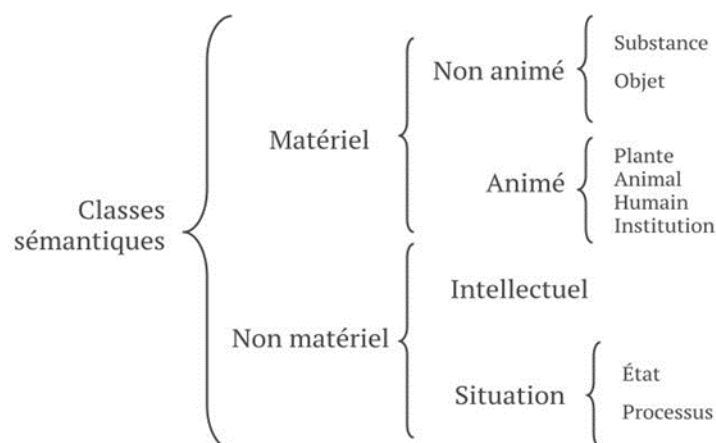


Figure 1. Classification sémantique utilisée pour PORTLEX (Domínguez, 2011).

Malgré les désaccords habituels, les grammairiens parient souvent pour un classement sémantique des substantifs plutôt général (Grevisse et Goosse, 2008 : 583-584 ou Riegel, Pellat et Rioul, 2009 : 323- 328), en établissant de grandes classes ou hyperclasses : animé vs inanimé, concret vs abstrait, comptable vs massif, individuel vs collectif etc., alors que d'autres (Charaudeau, 1992 ou Abeillé et Godard, 2021) proposent une classification plus fine et détaillée qui distingue des sous-classes spécifiques à l'intérieur des grandes classes.

Charaudeau (1992 : 17-23), par exemple, établit tout d'abord la différence entre êtres physiques et matériels, êtres de résultat d'un processus situé dans le temps et êtres ni physiques ni localisables dans le temps ; ceci fait, il opère un classement de ces classes d'êtres en fonction de classes d'appartenance, déterminées par des traits de caractère générique ou spécifique qu'il subdivise à la fois en plusieurs sous-classes assez précises et minutieuses.

D'autre part, dans la *Grande grammaire du français* (Abeillé et Godard, 2021 : 398-427), récemment publiée, après avoir fait mention des différentes hyperclasses (noms d'êtres humains ou animés, noms concrets, noms abstraits, noms de situation, noms de qualité ou de propriété et noms de qualité ou de proportion), on examine de près ces différentes catégories et on propose des distinctions et maintes subdivisions assez subtiles et particulières.

Finalement, même s'il ne s'agit pas d'une grammaire, il est tout à fait nécessaire pour cette question de mentionner G. Gross qui, dans son *Manuel d'analyse linguistique* paru en 2012, consacre tout un chapitre à ce problème. Afin de remédier à l'imprécision des hyperclasses sémantiques, ce linguiste et son école proposent une organisation sémantique du lexique basée sur la notion de « classe d'objet », qu'il définit comme

« des classes de substantifs, sémantiquement homogènes, qui déterminent une interprétation donnée d'un prédicat parmi d'autres possibles » (Gross 2012 : 75). De ce point de vue, puisqu'elles sont délimitées par les possibilités combinatoires au sein des prédicats, ces classes sémantiques sont beaucoup plus précises et donc appropriées au traitement automatique des langues (Blanco, 1997 : 373, 1999 : 7).

Malgré ces différentes propositions existantes pour le français, et aussi pour d'autres langues, dans le cadre des deux projets auxquels nous avons participé, MultiComb et MultiGenera, il a fallu concevoir un outil sur mesure (Domínguez, Valcárcel et Bardanca, 2021 ; Domínguez, 2022)¹⁰. Étant donné les particularités du travail à entreprendre, où une analyse plus fine des arguments était nécessaire, on ne pouvait plus faire appel aux hyperclasses sémantiques utilisées dans le dictionnaire PORTLEX (Domínguez, 2011). Il nous fallait donc un cadre d'analyse plus granulaire. Vu que notre objectif était avant tout d'organiser des paquets lexicaux dans une base de données multilingue, un classement sémantique de ceux-ci n'était pas prioritaire. Par conséquent, l'outil de catalogage finalement appliqué n'est pas basé sur des classes sémantiques ou des classes d'objets à proprement parler, mais sur des étiquettes sémantiques (Barrios, 2010 : 12, 25-26 ; Blanco, 2015 : 41-43). Par étiquette sémantique, nous entendons plutôt une paraphrase minimale des principaux traits sémantiques d'un lexème. Ce genre d'étiquettes relèvent donc des catégories descriptives plutôt que des catégories d'analyse (Polguère, 2003 : 50).

Dans les projets qui nous ont occupés, les étiquettes sémantiques étaient utilisées notamment comme une partie du système de codage des paquets lexicaux stockés dans la base de données¹¹. Pour mieux répondre à ce système de codage, les étiquettes sémantiques créées pendant l'analyse ont été organisées de façon hiérarchique en fonction, d'une part, des relations d'hyponymie et d'hyperonymie, et d'autre part, des contextes de combinatoire au sein des groupes nominaux¹². Cette organisation hiérarchique nous permettait, en même temps, d'éviter des polysémies inutiles et d'utiliser les mêmes étiquettes pour coder les paquets lexicaux des différents substantifs analysés

¹⁰ La hiérarchie d'étiquettes sémantiques utilisée peut être consultée en ligne : <http://portlex.usc.gal/ontologia>.

¹¹ En plus de l'étiquette sémantique, le nom de chaque paquet lexical comporte d'autres informations : langue, noyau du syntagme nominal et réalisation argumentale. Par exemple, dans le nom de fichier *fr_mort_estr1_animado_humano_familia* on indique qu'il s'agit d'un paquet lexical contenant des noms du champ sémantique des relations familiales (mère, père, oncle, grand-père etc.) sélectionnés pour générer la réalisation argumentale *de + déterminant + nom* de l'argument expérient (Abeillé et Godard, 2021 : 188) dépendant du prédicat nominal MORT (*estr1*). Comme on peut le voir, les étiquettes, ainsi que les noms des paquets, sont en espagnol, métalangue des projets en question.

¹² Le besoin d'avoir recours à un dispositif d'étiquetage plus granulaire et hiérarchique nous a menés à exclure des systèmes d'étiquetage sémantique déjà existants comme le USAS (UCREL Semantic Analysis System) ou les domaines sémantiques proposés par le SIL (Summer Institute of Linguistics) (Moe, 2003 ; Rayson *et al.*, 2004).

(Barrios, 2010 : 25-26)¹³. Compte tenu de tous ces éléments, la réutilisation de notre système d'étiquetage en dehors de ce cadre de travail spécifique risque bien d'être difficilement applicable.

1.3. Les sens du nom MORT

Le nom MORT est un vocable polysémique qui regroupe donc une grande variété de sens ou lexèmes. Certains dictionnaires de référence, dont le *TLFI*, le *LR* ou le *Lexis*, en distinguent deux grands groupes : d'une part, la cessation de la vie et, d'autre part, la destruction ou la fin de quelque chose.

Le premier groupe se rapporte à un patient animé, qui peut être soit un organisme vivant (humain, animal, plante, etc.) soit un organe de celui-ci (cerveau, poumons, etc.). Par exemple :

- (1) la mort de ma mère.
- (2) la mort cérébrale.

Tandis que le deuxième groupe fait référence à un patient non animé (objets, processus, etc.).

- (3) la mort de la voiture thermique.
- (4) la mort de son projet.

Pour notre recherche nous ne retenons que des acceptions du premier groupe (exemples 1, 2), à l'exception de celle concernant la « force intemporelle sentie comme une menace pour toute vie humaine » (*TLFI*). De même, nous avons exclu l'acception de MORT avec le sens de 'assassinat, meurtre', où la cause de la mort est remplacée par un agent humain, p.ex. : la mort de César par Brutus.

2. Méthodologie

La méthodologie que nous décrivons et appliquons dans ce travail a été conçue spécifiquement pour atteindre les objectifs de deux projets de recherche liés à la génération automatique de groupes nominaux : Multigenera et MultiComb. Même si l'espagnol était la métalangue de ces projets, le français et l'allemand étaient aussi des langues cibles et pour lesquelles on a développé non seulement une méthodologie pour extraire et annoter des données de corpus, mais aussi des outils sur mesure pour les analyser sémantiquement (Domínguez, Lindemann et Valcárcel, 2018 ; Domínguez, Solla et Valcárcel, 2019 ; Domínguez, 2022).

¹³ Par exemple, l'étiquette *animado_planta_arbol* est utilisée pour coder des paquets lexicaux associés à des arguments des prédicats d'ODEUR et de MORT. Cependant le contenu de ces paquets est différent : dans le cas du paquet associé à MORT (la mort du pin) on inclut les noms d'arbres en général, tandis que ceux associés à l'argument 'source' d'ODEUR (l'odeur du pin) n'incluent que des arbres qui dégagent une odeur caractéristique. La présence des noms des noyaux, ainsi que de codes associés aux réalisations argumentales, nous permet de les distinguer : *fr_mort_estr1_N1_animado_planta_arbol* face à *fr_odeur_estr3_N2_animado_planta_arbol*.

L'objectif principal de cette méthodologie est donc celui d'analyser la syntaxe et la sémantique des GN afin de pouvoir les générer automatiquement. Pour ce faire, les étapes parcourues pour chaque noyau nominal ont été les suivantes :

1. Étude du schéma argumental à partir de la consultation de différentes ressources existantes. Évidemment, le dictionnaire PORTLEX a constitué la source de départ, mais d'autres outils ont été utilisés tels que Framenet¹⁴, le *Lexique actif du français* (Mel'čuk et Polguère, 2007) ainsi que différentes ressources lexicographiques, dont le TLFi.
2. Recherche en corpus des occurrences argumentales. Nous avons utilisé le corpus *FrTenTen2017* (Jakubíček *et al.*, 2013) sur Sketch Engine (Kilgarriff, 2014).
3. Dépouillement des occurrences obtenues à plusieurs niveaux (voir ci-dessous).
4. Annotation sémantique des occurrences retenues pour chaque réalisation argumentale. Tel que l'on vient d'expliquer dans la section précédente, nous avons utilisé un système d'étiquetage conçu à cet effet dans le cadre de nos analyses.

Pour obtenir des données précises sur le corpus, nous avons utilisé le langage CQL (Corpus Query Language) (Lexical Computing, 2023a), qui nous a permis de faire des recherches précises en fonction de différents critères. Dans ce cas, nous avons décidé d'extraire les données uniquement pour la première position argumentale des prédicats nominaux, c'est-à-dire celle qui est contiguë à leur noyau. Ce choix a dû être fait pour simplifier les requêtes et pour éviter d'obtenir des résultats ambigus ou non pertinents. En effet, la considération d'autres positions argumentales dans nos recherches aurait impliqué de définir des expressions régulières plus complexes et de dépouiller un nombre beaucoup plus important de résultats¹⁵. Par ailleurs, nous avons décidé de ne pas inclure de recherches avec des déterminants au pluriel (p.ex., les morts, ces morts) car cela pouvait être interprété de différentes manières : « les morts » peut signifier 'le nombre de décès' ou 'les personnes décédées'. Or, dans notre cas, nous ne nous intéressons qu'au premier sens, c'est-à-dire au nombre de décès. Finalement, pour certains arguments nous avons limité nos recherches à des réalisations dont les résultats nous fournissaient déjà les données nécessaires aux objectifs fixés par les projets de recherche en question¹⁶.

À partir des expressions CQL formulées (voir annexe), les données obtenues en format KWIC (keyword in context) (Lexical Computing, 2023b) comprennent les

¹⁴ Nous avons consulté aussi bien Framenet pour l'anglais (Ruppenhofer *et al.*, 2016) que la version développée pour le français (Djemaa *et al.*, 2016).

¹⁵ Il convient toutefois de rappeler ici que cette extraction de données a été réalisée pour rechercher des prototypes lexicaux et sémantiques dans le cadre de deux projets de recherche, et non pour mener une analyse exhaustive des prédicats nominaux.

¹⁶ Il s'agit concrètement de la réalisation *déterminant + mort + par + déterminant + nom* (p.ex. : sa mort par un cancer des poumons inopérable).

mots (prépositions, déterminants, noms et adjectifs) qui constituent le contexte du noyau étudié (mort), c'est-à-dire, les arguments à analyser, par exemple : la mort de sa mère, la mort par cancer, la mort cellulaire¹⁷. La fréquence de chaque KWIC a été obtenue et finalement, seuls les éléments dont la fréquence est égale ou supérieure à 5 occurrences sont retenus pour le dépouillement et l'étude. En tout cas, les analyses ont été limitées à 1000 KWIC par réalisation argumentale.

Une fois les données à traiter délimitées, nous passons au dépouillement des données obtenues en trois phases : un premier triage pour ne retenir que les occurrences argumentales (p.ex., « la mort par hasard » versus « la mort par cancer »), un deuxième pour retenir les occurrences correspondant au sens visé (p.ex., « la mort de la révolution » versus « la mort de son oncle ») et un troisième pour classer les occurrences formellement similaires en fonction des arguments (expérient ou cause, p.ex. : « la mort de Pierre » versus « la mort de soif »).

Ensuite, nous réalisons une lemmatisation des occurrences dépouillées¹⁸ pour parvenir à l'annotation sémantique de chaque lemme résultant¹⁹. Nous obtenons ainsi des données sur la fréquence de chaque lexème annoté dans une réalisation argumentale donnée. Comme nous le verrons par la suite, cela nous permet non seulement de repérer quels sont les lexèmes les plus fréquents dans chaque étiquette sémantique, c'est-à-dire d'identifier des prototypes lexicaux, mais aussi de calculer la fréquence des différentes étiquettes sémantiques associées à une réalisation argumentale. Nous pouvons ainsi identifier des prototypes sémantiques pour chaque argument, c'est-à-dire les étiquettes sémantiques qui apparaissent le plus fréquemment dans celui-ci et qui le caractérisent donc le plus. Et finalement, nous pouvons calculer aussi quelles sont les réalisations argumentales les plus fréquentes dans un prédicat nominal, au moins en première position.

¹⁷ Afin de simplifier les recherches, dans les expressions CQL utilisées on ne considère que le premier mot situé après les prépositions et leurs contractions. Cela veut dire que les KWIC obtenus ne montrent pas le noyau des arguments quand des adjectifs ou des noms composés sans trait d'union (p.ex., jeune femme, jolie fille etc.) se trouvent dans cette position, p.ex. : la mort de la jeune [femme], la mort de cette jolie [fille]. Par conséquent, ces cas-là sont exclus de nos analyses car le manque de contexte nous empêche d'attribuer une étiquette sémantique. Pour la même raison, nous avons rejeté aussi des occurrences sûrement argumentales comme « la mort de la moitié » ou « la mort de la plupart ».

¹⁸ Par exemple, sous un même lemme MAÎTRESSE on inclut les données de fréquence des KWIC suivants : la mort de la maîtresse, d'une maîtresse, des maîtresses, de ces maîtresses, de ses maîtresses, etc.

¹⁹ Le processus d'annotation sémantique implique souvent de traiter des cas de désambiguïsation. Par exemple, dans le KWIC « la mort de la femme », il est impossible de savoir, sans vérifier exemple par exemple, si le mot *femme* est synonyme d'épouse (la femme de Louis) ou s'il fait référence au genre d'une personne (la femme de la maison d'à côté). Dans ce cas, on attribue l'étiquette la plus générale : être humain féminin. Cependant, dans le KWIC « la mort de sa femme » le possessif indique clairement qu'il s'agit du sens 'épouse'.

3. Analyse des données

3.1. Arguments

Dans les acceptions de MORT que nous avons retenues pour notre recherche, on peut identifier deux arguments, un expérient et une cause²⁰. Ces deux arguments sont exigés par le prédicat, mais leur coprésence n'est pas obligatoire :

- (1) La mort de mon oncle par cancer est difficile à accepter
- (2) La mort de mon oncle est difficile à accepter
- (3) La mort par cancer est difficile à accepter.
- (4) La mort est difficile à accepter.

Comme nous l'avons vu dans la section précédente, l'exemple (4) correspondrait déjà à une autre acception, car ni la cause ni l'expérient ne sont exprimés.

3.1.1. Expérient

Le prédicat MORT comporte un expérient animé qui éprouve la cessation de la vie. Dans ce sens, nous suivons Abeillé et Godard (2021 : 188), ainsi que Riegel *et al.* (2009 : 237), qui distinguent un rôle d'expérient ou expérienceur différent du rôle de patient. Cette distinction tient à la présence ou non d'un agent, normalement exigé par le patient (Charaudeau, 1992 : 383)²¹.

L'expérient peut être omis, mais dans ce cas on évoque une sorte de généralisation²² :

- (5) Les chercheurs ont beaucoup travaillé sur la mort par grippe [de n'importe qui].

Deux réalisations formelles sont possibles pour exprimer cet argument :

- a. Un complément déterminatif précédé par la préposition *de*, avec un déterminant optionnel :
 - (6) La mort de ma mère.
 - (7) La mort de soldats français.
 - (8) La mort de Jacques.
- b. Une réalisation adjectivale :
 - (9) La mort cellulaire

²⁰ Pour plus de clarté, nous adoptons dans cette description les dénominations habituelles dans les grammaires du français (Abeillé et Godard, 2021 ; Charaudeau, 1992 ; Riegel, Pellat et Rioul, 2009). Cependant, dans le cadre des projets où nos recherches se sont développées, on a eu recours à une autre nomenclature, créée *ad hoc* et basée sur les travaux de M^a José Domínguez (2011).

²¹ Tel que nous l'avons déjà commenté, le prédicat MORT comportant un patient et un agent humain correspond à une autre acception, dont il n'est pas question dans ce travail (voir 1.3).

²² Cette omission nous fait penser aux emplois dits absolus de certains verbes comme MANGER ou LIRE, p.ex., cf. « il mange de la viande chez lui » par opposition à « il mange chez lui ».

3.1.2. Cause

Tout en étant un argument circonstanciel (Charaudeau, 1992 : 386), nous avons inclus la cause dans notre analyse parce que plusieurs courants linguistiques considèrent les composantes circonstancielles comme des actants, de même que certaines grammaires de référence (Abeillé et Godard, 2021 : 188-189 ; Riegel, Pellat et Rioul, 2009 : 238). En plus, dans le cas particulier du prédicat qui nous occupe, la cause est très fréquemment explicitée et même obligatoire quand l'expérient est sous-entendu²³.

Dans le cas de cet argument, trois réalisations formelles sont disponibles :

- a. Un complément déterminatif précédé par la préposition *de*, sans déterminant :

(10) La mort de soif.

- b. Un complément déterminatif introduit par la préposition *par*, avec un déterminant optionnel²⁴ :

(11) La mort par noyade.

(12) La mort par une noyade accidentelle.

- c. Un adjectif :

(13) La mort atomique

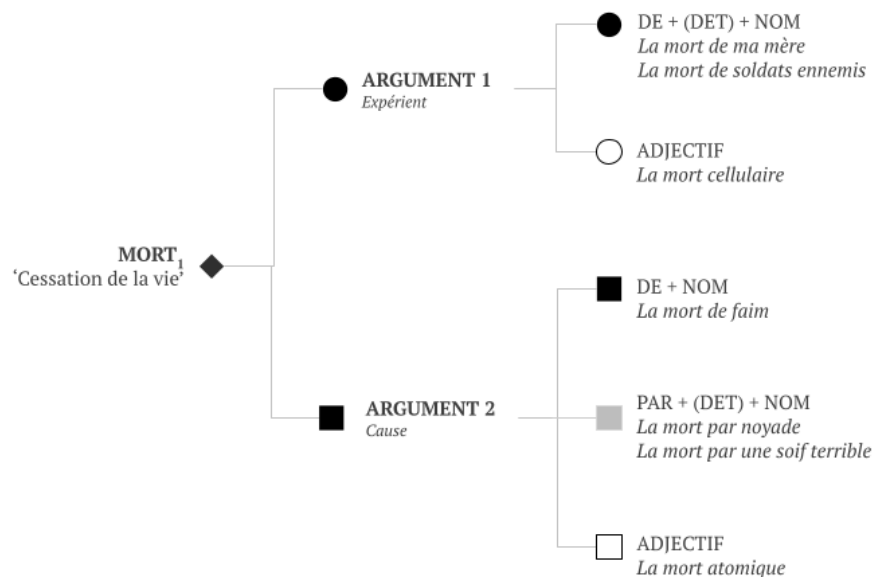


Figure 2. Schéma argumental de MORT₁.

²³ Cette contrainte que nous observons dans ce prédicat MORT pourrait questionner en quelque sorte le statut non essentiel de l'argument cause, étant donné que son absence entraîne un changement de sens (cf. exemples 3 et 4).

²⁴ La détermination dans cette réalisation est obligatoire dès que le noyau de l'argument est accompagné d'un modifieur (adjectif, relatif etc.).

La réalisation la plus fréquente est la complémentation au moyen de la préposition *par*, tandis que le recours à l'adjectivation est plutôt rare²⁵. En ce qui concerne la première réalisation, un complément introduit par la préposition *de*, il paraît qu'une restriction sémantique est à appliquer. Tous les cas que nous avons repérés comportent des noms renvoyant à un état physique (la mort de faim, de soif, d'épuisement etc.) ou psychique (la mort de chagrin, de peur etc.). En général, les noms présentant ce trait sémantique sont rarement utilisés dans des complémentations avec la préposition *par* (la mort par chagrin, par épuisement etc.). Cependant, dès qu'on ajoute un modifieur, la construction avec *par* suivi d'un déterminant est de règle :

- (14) La mort de chagrin. (rare)
- (15) La mort par chagrin. (rare)
- (16) La mort par chagrin d'amour.

3.2. Combinatoire

Les deux arguments de MORT sont susceptibles d'apparaître combinés en occupant les positions intérieure ou extérieure²⁶.

- (16) La mort des animaux par étouffement.
- (17) La mort par étouffement des animaux.

Cependant il y a certaines restrictions dont il faut tenir compte :

- a. Certaines réalisations sont figées dans une seule position : c'est le cas de la réalisation *de + nom* de l'argument causal, qui doit toujours être placé en position intérieure.
 - (18) La mort de faim de son frère.
 - (19) *La mort de son frère de faim.
- b. Toujours en position intérieure, la réalisation *de + nom* de l'argument expérient ne peut être combinée qu'avec la réalisation *par + nom* de l'argument causal. Cependant, en position extérieure, cette réalisation expérientielle n'affiche pas de restrictions dans ce sens et peut accompagner tout l'éventail de réalisations formelles de l'argument causal.
 - (19) La mort de Julie par accident.
 - (20) *La mort de Julie [accidentelle ~ de faim].
 - (21) La mort par accident de Julie.
 - (22) La mort accidentelle de Julie.

²⁵ Dans nos recherches en corpus, nous n'avons pas pu trouver plus de trois exemples de cette réalisation, dont deux (la mort accidentelle, la mort naturelle) ne nous semblent pas évidents.

²⁶ Nous utilisons ici les termes « position intérieure » et « position extérieure » avec un sens strictement formel, ce qui nous permet d'éviter des confusions avec d'autres concepts souvent utilisés dans la description des schémas argumentaux, tels que « position centrale » ou « position périphérique », et qui renvoient à des notions plus sémantiques.

- (23) La mort de faim de Julie.
- c. Parallèlement, quand la réalisation causale *par + nom* occupe la position intérieure, la réalisation expérientielle *de + nom* est de mise.
- (24) La mort par accident des parents.
- (25) *La mort par accident parentale.
- d. Quant aux réalisations adjectivales, c'est l'argument causal qui semble présenter le plus de particularités. En position intérieure, cette réalisation exige une suite prépositionnelle *de + nom*, alors qu'en position extérieure elle semble exiger un autre adjectif expérient²⁷.
- (26) La mort cardiaque de ce patient.
- (27) Une mort royale accidentelle. (rare)

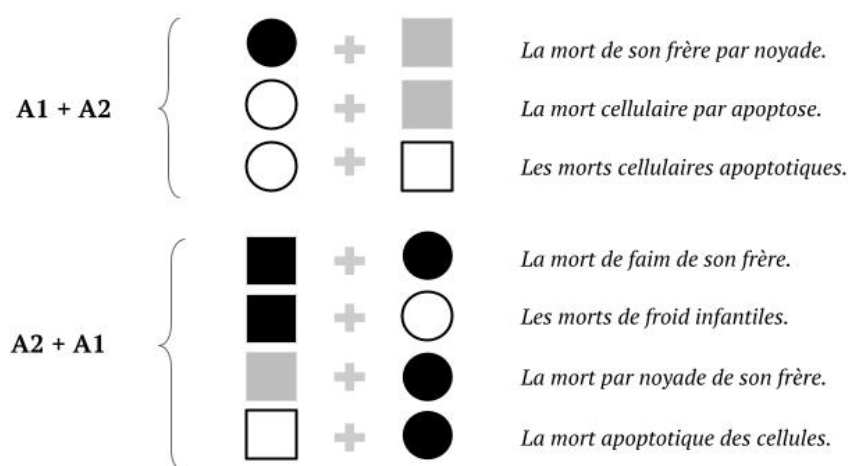


Figure 3. Schémas combinatoires des arguments de MORT₁ par réalisation formelle.

3.3. Fréquence des arguments en position intérieure

Quant à la fréquence des arguments en position intérieure, les données obtenues dans nos projets sont claires et nettes pour le cas de MORT : c'est l'argument expérient qui apparaît le plus souvent à côté du noyau de ce groupe nominal. Il semble évident que pour les locuteurs il est prioritaire de préciser avant qui meurt plutôt que de spécifier la cause de la mort. En fait, le nombre de lemmes retenus (95) pour l'argument cause, ainsi que leur fréquence moyenne (77,7), sont nettement inférieurs aux chiffres affichés pour l'argument expérient (635 et 274,4 respectivement).

²⁷ Il faut quand même signaler que les exemples obtenus dans les corpus pour cette réalisation causale en position extérieure ont été très peu nombreux, ce qui nous empêche de tirer des conclusions probantes sur ce schéma argumental.

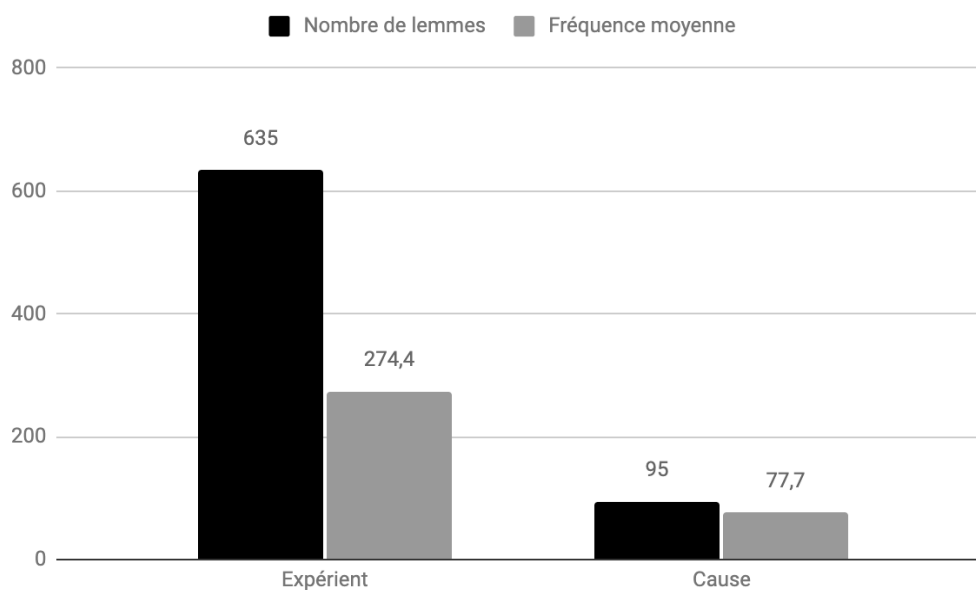


Figure 4. Fréquence moyenne des cooccurrences des arguments de MORT₁.

Si nous examinons ces chiffres au niveau des réalisations argumentales, nous constatons que les réalisations adjectivales sont très fréquentes. Pour l'argument expérient, la fréquence moyenne de cette réalisation égale celle des occurrences avec préposition, alors que pour l'argument causal elle est nettement plus présente dans le corpus. Ces chiffres de fréquences contrastent avec ceux affichés pour le nombre total de lemmes de ces réalisations (Figure 4). Malgré leur fréquence, on voit que le nombre de lemmes des réalisations adjectivales est très réduit par rapport à celui des occurrences prépositionnelles : 23 pour l'argument expérient et 13 pour l'argument causal. Ces écarts entre fréquence et nombre absolu suggèrent à la fois un fort degré de rentabilité pour les lemmes de ces réalisations et une certaine spécialisation sémantique. Il s'agit en effet d'exemples très courants dans l'usage, p.ex. : mort cérébrale, mort cellulaire, mort cardiaque, mort accidentelle, mort violente, etc. Nous pensons d'ailleurs que dans certains cas, il peut s'agir en réalité de réalisations adjectivales de compléments circonstanciels, mais leur fréquence élevée dans le corpus utilisé rend leur désambiguïsation irréalisable²⁸.

²⁸ Des études comme celle de Gemma Rigau (1999) pour l'espagnol montrent pourquoi les adjectifs relationnels peuvent être des réalisations argumentales ou circonstanciels, tout en remarquant que le contexte est, comme toujours, un facteur déterminant dans leur classification.

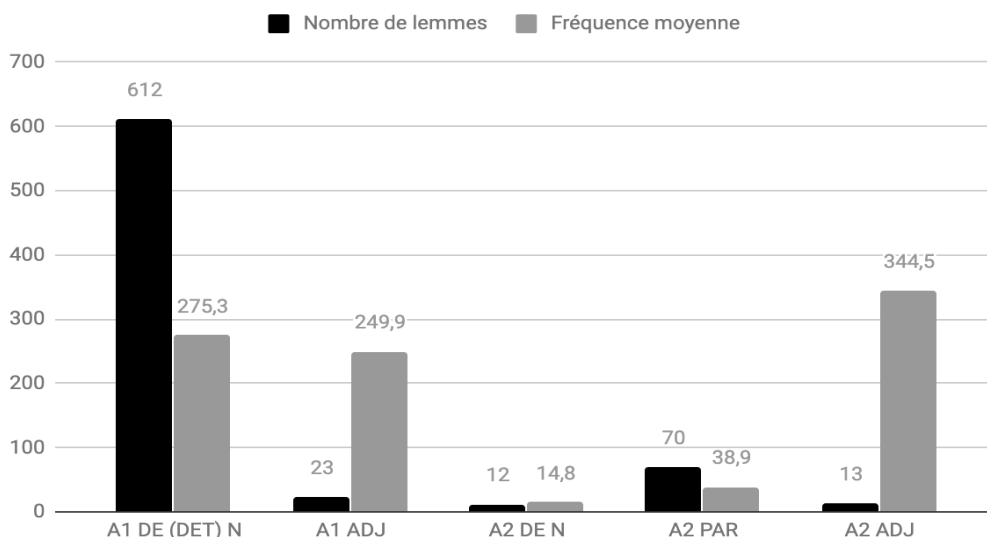


Figure 5. Fréquence moyenne des cooccurrences des arguments de MORT par réalisation formelle.

3.4. Sémantisme des arguments

L’annotation sémantique des lemmes retenus pour chaque argument fait ressortir certains comportements que nous tenons à commenter ici. Tout d’abord, on peut constater que c’est l’argument expérient qui regroupe la plupart des étiquettes sémantiques.

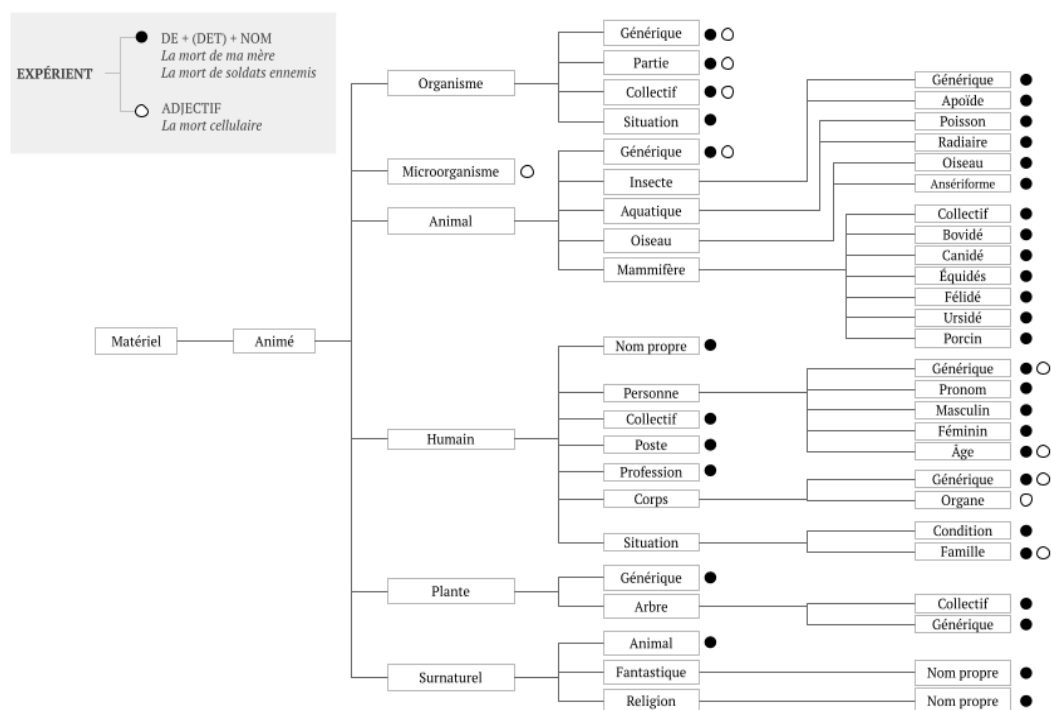


Figure 6. Annotation sémantique des cooccurrences de l’argument expérient par réalisation formelle.

En effet, c'est aussi l'argument qui présente le plus grand nombre de lemmes et la fréquence la plus élevée. Au total, nous avons identifié 55 étiquettes différentes²⁹, dont la plupart correspondent à la réalisation prépositionnelle.

Quant à l'argument causal, les étiquettes sont beaucoup moins nombreuses, mais cela ne devrait pas pour autant nous amener à penser que cet argument est moins diversifié sémantiquement que l'expérient : au total, nous avons identifié 29 étiquettes sémantiques à plusieurs niveaux³⁰. Cependant, à la différence de l'argument expérient, cet argument présente déjà dès le premier niveau une spécification sémantique entre les causes matérielles et non matérielles de la mort (p.ex. « mort par bombe » face à « mort par cancer »). La diversification sémantique de l'argument expérient se produit à niveaux inférieurs de notre hiérarchie, lorsqu'il faut spécifier le type et la condition de l'entité animée qui éprouve la mort (p.ex. la mort de ton père, la mort des poissons, la mort du palmier, etc.).

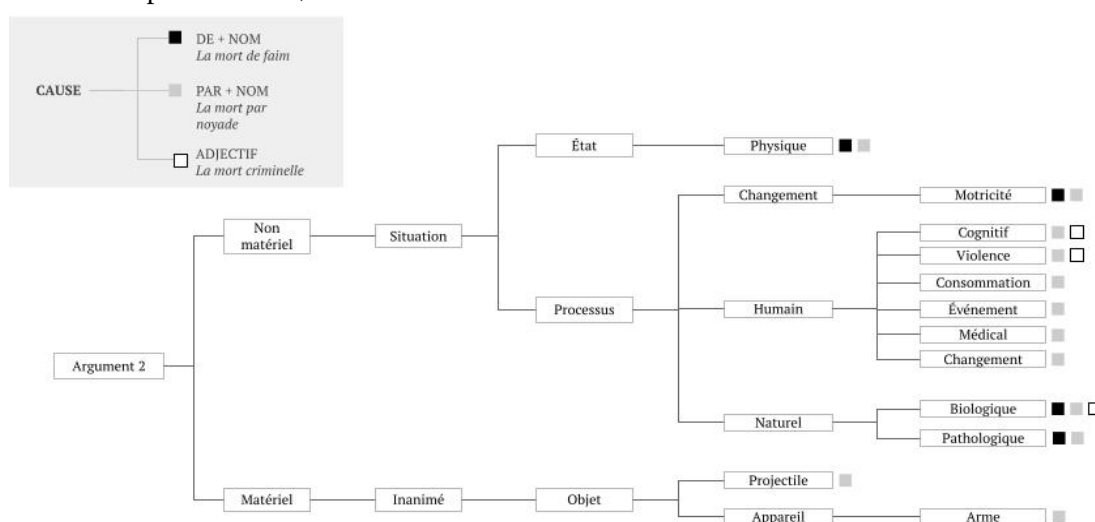


Figure 7. Annotation sémantique des cooccurrences de l'argument cause par réalisation formelle.

Au niveau des réalisations argumentales, l'annotation sémantique dégage aussi des résultats probants. D'une part, nous observons des réalisations qui recouvrent presque toutes les étiquettes sémantiques associées à un rôle argumental. C'est le cas des réalisations prépositionnelles *de + déterminant + nom* de l'argument expérient et *par + nom* de l'argument causal. D'autre part, certaines réalisations sont généralement plus rattachées à des traits sémantiques spécifiques. Dans les résultats obtenus, cela ressort clairement dans les réalisations adjectivales des deux arguments, mais surtout dans la réalisation prépositionnelle *de + nom* de l'argument causal (mort de faim, mort de

²⁹ En termes de hiérarchie, nous avons 6 étiquettes sémantiques de troisième niveau, 21 de quatrième niveau et déjà 26 au cinquième niveau.

³⁰ Nous avons noté 2 étiquettes de premier niveau et 2 de deuxième niveau, 3 étiquettes de troisième niveau, 6 de quatrième niveau et 10 de cinquième niveau.

soif), où l'étiquette *non_matériel_situation_état physique* est clairement prédominante. Même si les occurrences de ces réalisations avec d'autres étiquettes sont possibles, nous avons donc affaire à une spécialisation sémantique assez marquée, comme le suggèrent déjà les données de fréquence.

Cette spécialisation sémantique nous fait penser à l'existence de prototypes sémantiques non seulement pour chaque argument, mais aussi pour chaque réalisation (Kleiber, 1991). De même, nous avons pu identifier des prototypes lexicaux, c'est-à-dire les lexèmes les plus fréquents pour une étiquette sémantique au niveau des réalisations argumentales. Par exemple, dans les réalisations *expérient de (+ déterminant) + nom* les cinq lexèmes les plus fréquents affichent l'étiquette sémantique *matériel_animé_humain_situation_famille* : père (16847 occurrences), mère (9833), parents (6524), fils-fille (5312), frère (4575). Ces cinq lexèmes constituent à eux seuls 24,81% du total des occurrences. Cela veut dire que presque un quart des cas dépouillés pour ces deux réalisations correspond seulement à cinq lexèmes (0,74% du total) rattachés à la même étiquette sémantique.

Des prototypes sémantiques et lexicaux ont été repérés pour les réalisations mais, malheureusement, il s'avère très compliqué de présenter ces données si détaillées³¹, raison pour laquelle nous ne les avons pas incluses dans cette étude à caractère plutôt général.

4. Conclusions

Nous touchons à la fin de notre travail, où nous avons décrit une méthodologie précise pour analyser des prédicats nominaux, développée dans le cadre particulier des projets MultiGenera et MultiComb. Plus précisément, nous avons montré en détail son application à un cas concret, le prédicat MORT. Notre analyse n'a pas épargné non plus des réflexions sur la complexité du GN et surtout la question souvent épineuse de la polysémie des vocables.

La première constatation à retenir est la grande quantité de données recueillies dans le corpus, qui exigent un dépouillement sélectif avant d'envisager le calcul de la fréquence des occurrences et leur annotation syntaxico-sémantique. Il faut rappeler que les données traitées dans ce travail sont partielles du moment que nous visions initialement des objectifs autres que l'analyse exhaustive des prédicats nominaux, c'est-à-dire l'obtention de prototypes lexico-sémantiques pour la génération automatique de GN. Pourtant, il nous semble évident qu'une application plus large de cette méthodologie serait possible afin de couvrir tous les schémas argumentaux d'un prédicat donné.

D'ailleurs, cette analyse grammaticale combinant l'obtention de données de fréquence en corpus avec leur annotation sémantique a été extrêmement utile pour

³¹ Par exemple, si nous avons à présenter les cinq principaux prototypes lexicaux pour chacune des 84 étiquettes sémantiques identifiées, on devrait prévoir une liste de 420 lexèmes annotés, tâche qui devrait faire l'objet d'une analyse à part.

approfondir nos connaissances sur le fonctionnement des prédicats nominaux. En fait, les résultats observés pour le cas de *MORT* nous ont permis de relever une série de questions pertinentes à cet égard :

1. Les données de combinatoire semblent montrer que, au sein du prédicat *MORT*, l'argument expérient est plus central que l'argument cause, étant donné que celui-ci apparaît beaucoup plus souvent en position intérieure, c'est-à-dire à proximité du noyau. Le degré de centralité d'un argument au sein d'un prédicat a été mis en relation avec la contiguïté par rapport au noyau (Lazard, 1994 : 14-15, 1998: 18-21 ; Lehmann, 2015 : 1549). Compte tenu de ce que nous avons pu observer pour le cas de *MORT*, il semble que les prédicats nominaux se comportent à cet égard d'une manière similaire aux prédicats verbaux. Cependant, il faudrait explorer d'autres prédicats nominaux afin de mieux cerner l'application du concept de centralité argumentale aux GN, qui présentent de nombreuses particularités.
2. Grâce à la combinaison des données de fréquence avec l'annotation sémantique, nous avons pu constater que dans un même argument les réalisations présentent une spécialisation sémantique, c'est-à-dire il y a certaines étiquettes qui sont nettement plus fréquentes que d'autres, ce qui nous a permis de relever des prototypes. C'est précisément sur ces prototypes sémantiques au niveau des réalisations que portait notre recherche au sein des projets *Multigenera* et *Mul-tiComb* (Domínguez, Valcárcel et Solla, 2019).
3. Cette démarche nous a permis d'identifier, non seulement des prototypes sémantiques, mais aussi des prototypes lexicaux pour chaque étiquette. Cela veut dire que pour chaque réalisation il existe aussi des lexèmes prototypiques associés au sémantisme de son argument.

Malgré les défauts qu'elle puisse avoir, la méthodologie appliquée pour les projets mentionnés ci-dessus et présentée dans ce travail nous permet de mieux cerner des aspects cruciaux pour étudier les prédicats nominaux. En fin de compte, ce type d'analyse revient à étudier les mots dans leur contexte, ce qui est fréquemment défendu par les spécialistes en didactique des langues (Laufer et Nation 2012 ; Nation 2001, 2005). A cet égard, il nous semble pertinent de rappeler l'importance pédagogique des exemples dans l'enseignement des langues ainsi qu'en lexicographie. Précisément, des questions comme la centralité des arguments ou l'identification de prototypes sémantiques et lexicaux au sein des GN pourraient s'avérer essentielles lors de la sélection ou l'élaboration d'exemples dans les grammaires et dans les dictionnaires.

Vu l'utilité que nous attachons à la méthodologie retenue pour cette étude, il nous paraît nécessaire de poursuivre les analyses pour inclure des arguments nominaux dans des positions autres que l'intérieure ou pour tester l'efficacité d'autres systèmes d'annotation sémantique. Enfin, il faudrait explorer la validité de cette méthodologie pour l'étude d'autres types de prédicats, notamment pour les prédicats verbaux.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ABEILLÉ, Anne & Danièle GODARD [dirs.] (2021) : La grande grammaire du français. Arles, Actes Sud.
- BARRIOS, Ma Auxiliadora (2010) : « El dominio de las funciones léxicas en el marco de la Teoría Sentido-Texto ». *Estudios de Lingüística del español (ELiEs)*, 30. URL : <http://elies.rediris.es/elies30/index30.html>.
- BLANCO, Xavier (1997) : « De las clases de objetos a las clases de predicados ». *Verba*, 24, 371-385.
- BLANCO, Xavier (1999) : *Lexicographie bilingue français-espagnol et classes d'objets*. Barcelone, Universitat Autònoma de Barcelona.
- BLANCO, Xavier (2015) : « Structurer le lexique de l'espagnol moyennant des classes sémantiques ». *Verbum*, 6, 41-52.
- CHARAUDEAU, Patrick (1992) : *Grammaire du sens et de l'expression*. Paris, Hachette-Éducation.
- DJEMAA, Marianne *et al.*, (2016) : « Corpus annotation within the French FrameNet : a domain-by-domain methodology », in N. Calzolari, K. Choukri, T. Declerck & A. Moreno (coords.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, European Language Resources Association (ELRA), 3794-3801.
- DOMÍNGUEZ, María José (2011) : *Kontrastive Grammatik und Lexikographie: spa-nisch-deutsches Wörterbuch zur Valenz des Nomens*. Munich, Iudicium.
- DOMÍNGUEZ, María José; Carlos VALCÁRCCEL & David LINDEMANN (2018) : « Multilingual Generation of Noun Valency Patterns for Extracting Syntactic-Semantical Knowledge from Corpora (MultiGenera) », in J. Čibej, V. Gorjanc, I. Kosem & S. Krek (eds.), *Lexicography in Global Contexts: Proceedings of the 18th EURALEX International Congress. Ljubljana*, Ljubljana University Press, 847-854.
- DOMÍNGUEZ, María José; Miguel Ángel SOLLA & Carlos VALCÁRCCEL (2019) : « Resources interoperability: exploiting lexicographic data to automatically generate dictionary examples », in Kosem, Iztok *et al.* (eds.), *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference. Brno, Lexical Computing CZ*, 51-71. Disponible sur: https://ellex.link/ellex2019/wp-content/uploads/2019/09/eLex_2019_4.pdf
- DOMÍNGUEZ, María José & Carlos VALCÁRCCEL (2020) : « PORTLEX as a multilingual and cross-lingual online dictionary », in M.J. Domínguez, M. Mirazo & C. Valcárcel (eds.), *Studies on multilingual lexicography*. Berlin, De Gruyter, 135-158
- DOMÍNGUEZ, María José; Carlos VALCÁRCCEL & Daniel BARDANCA (2021) : *Ontología léxica*. Santiago de Compostela. Disponible sur: <http://portlex.usc.gal/ontologia/>.
- DOMÍNGUEZ, María José (2022) : « Contribución de la semántica combinatoria al desarrollo de herramientas digitales multilingües ». *Círculo de Lingüística Aplicada a la Comunica-*

- ción, 90, 171-188. Disponible sur: <https://revistas.ucm.es/index.php/CLAC/article/view/73849/4564456560240GREVISSE>, Maurice & André GOOSSE (2008): *Le bon usage. Grammaire française*. Bruxelles, De Boeck-Duculot, 14^e éd.
- GROSS, Gaston (1991) : « Syntaxe du complément de nom ». *Linguisticae Investigationes*, 15 : 2, 255-284.
- GROSS, Gaston (2008) : « Les classes d'objets ». *Lalies*, 28, 111-165.
- GROSS, Gaston (2012) : *Manuel d'analyse linguistique*. Villeneuve d'Ascq, Presses Universitaires du Septentrion.
- GÓMEZ, Francisco Xavier & Miguel Ángel SOLLA (2018) : « Building the galician wordnet: methods and applications ». *Language Resources and Evaluation*, 52 : 1, 317-339. URL : <https://doi.org/10.1007/s10579-017-9408-5>.
- JAKUBÍČEK, Miloš *et al.* (2013) : « The Tenten Corpus Family », in A. Hardie et R. Love (eds.), *Proceedings of the 7th International Corpus Linguistics Conference CL*. Lancaster, Lancaster University, 125-127. URL : https://www.sketchengine.eu/wp-content/uploads/The_TenTen_Corpus_2013.pdf
- KILGARRIFF, Adam *et al.* (2012) : « The Sketch Engine: ten years on ». *Lexicography*, 1, 7-36.
- KLEIBER, Georges (1991) : *La sémantique du Prototype, catégories et sens lexical*. Paris, Presses Universitaires de France.
- LAUFER, Batia & Paul NATION (2012) : « Vocabulary », in S.M. Gass et A. Mackey (eds.), *The Routledge Handbook of Second Language Acquisition*. Londres / New York, Routledge, 163-176.
- LAZARD, Gilbert (1994) : *L'actance*. Paris, Presses Universitaires de France.
- LAZARD, Gilbert (1988) : « Définition des actants dans les langues européennes », in J. Feuillet (éd.), *Actance et valence dans les langues de l'Europe*. Berlin, Mouton de Gruyter, 11-146.
- LEHMANN, Christian (2015) : « Situation types, valency frames and operations », in A. Malchukov & B. Comrie (eds.), *Valency classes in the world's languages*. Berlin, Mouton de Gruyter, 1427-1480.
- LEXICAL COMPUTING (2023a) : « CQL - Corpus Query Language », *Sketch Engine*. URL : <https://www.sketchengine.eu/documentation/corpus-querying>
- LEXICAL COMPUTING (2023b) : « Concordance - a tool to search corpus », *Sketch Engine*. URL : <https://www.sketchengine.eu/guide/concordance-a-tool-to-search-a-corpus/>
- MEL'ČUK, Igor & Alain POLGUÈRE (2007) : *Lexique actif du français: l'apprentissage du vocabulaire fondé sur 20000 dérivations sémantiques et collocations du français*. Bruxelles, De Boeck.
- MOE, Ronald (2003) : « Compiling dictionaries using semantic domains ». *Lexicos*, 13, 215-223. URL : <https://www.ajol.info/index.php/lex/article/view/51389/40043>
- NATION, Paul (2001) : *Learning vocabulary in another language*. Cambridge, Cambridge University Press.

- NATION, Paul (2005) : « Teaching and learning vocabulary », in E. Hinkel (ed.), *Handbook of research in second language teaching and learning*. Mahwah / New Jersey, Erlbaum, 581-595.
- PARIOLLAUD, Fanelly (2008) : *Verbes: questions de sémantique*. Rennes, Presses Universitaires de Rennes.
- POLGUÈRE, Alain (2003) : « Étiquetage sémantique des lexies dans la base de données DiCo ». *Traitement Automatique des Langues*, 44 : 2, 39-68.
- RAYSON, Paul *et al.* (2004) : « The UCREL semantic analysis system », in T. Mitamura *et alii* (eds.), *Proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation (LREC 2004)*. Lisbonne, European Language Resources Association, 7-12. URL: http://www.lancaster.ac.uk/staff/rayson/publications/usas_lrec04ws.pdf
- RIEGEL Martin; Jean-Christophe PELLAT & René RIOUL (2009) : *Grammaire méthodique du français*, Paris, Presses Universitaires de France, 4^e éd.
- RIGAU, Gemma (1999) : « La estructura del sintagma nominal: los modificadores del nombre », in I. Bosque & V. Demonte (dirs.), *Gramática descriptiva de la lengua española*. Madrid, Espasa Calpe, 311-362.
- RUPPENHOFER, Josef *et al.* (2016) : *FrameNet II: Extended Theory and Practice*. Berkeley, International Computer Science Institute. URL : <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>.
- STAGE, Lilian (1994) : « La valence des noms en français », in M. Herslund (ed.), *Noun Phrase Structures*. Copenhagen, Samfundslitteratur, 93-131.
- STAGE, Lilian (1997) : « La transposition des actants dans le syntagme nominal. Étude sur la nominalisation nucléaire et l'emploi des prépositions ». *Revue Romane*, 32 : 1, 51-86.

DICIONNAIRES

LEXIS : *Le Lexis. Dictionnaire érudit de la langue française*. Paris, Éditions Larousse.

LR : *Le Robert. Dico en ligne*. Paris, Le Robert. Disponible sur: <https://dictionnaire.lerobert.com/>

PORTLEX : *PORTLEX. Diccionario multilingüe de la frase nominal*. Santiago de Compostela, Universidade de Santiago de Compostela. Disponible sur: <http://portlex.usc.gal/diccionario/>

TLFi : *Trésor de la langue Française informatisé*. Nancy, ATILF-CNRS et Université de Lorraine. Disponible sur: <http://www.atilf.fr/tlfi>

EXPRESSIONS CQL FORMULÉES

- *Déterminant + MORT + de + nom*
- *Déterminant + MORT + de + déterminant + nom*
- ([tag="DET:POS"|word="la|une|cette"])([word="mort"] [word="de|du|des"] [tag="NOM|NAM"])([word="mort"] [word="de|d'"] [tag="DET:ART|DET:POS"|word="ce|cet|cette|ces"] [tag="NOM"])([word="mort"] [word="

"de"]

[tag="DET:ART"][tag="NOM|NAM"])([word="mort"][word="de|d"])[tag="NOM|NAM"])

– *Déterminant + MORT + par + nom*

– ([tag="DET:POS"|word="la|une|cette"])[word="mort"] [word="par"][tag="NOM|NAM"]

– *Déterminant + MORT + adjectif*

– [word="la|une|cette|ma|ta|sa|notre|votre|leur|aucune|quelle"][word="mort"][tag="ADJ"]