

ARTICLE

Pour une classification sémantique des noms en français appuyée sur des tests linguistiques

Pauline Haas¹, Lucie Barque², Richard Huyghe³ and Delphine Tributou⁴

¹Université Sorbonne Paris Nord & Lattice (UMR 8094, CNRS, ENS & Paris Sorbonne Nouvelle; PSL & USPC), ²Université Sorbonne Paris Nord & LLF (UMR 7110 CNRS et Université Paris Cité),

³Université de Fribourg and ⁴Université de Lille & STL (UMR 8163 - CNRS)

Corresponding author: pauline.haas@univ-paris13.fr

(Received 27 July 2021; revised 22 April 2022; accepted 04 May 2022; first published online 28 July 2022)

Résumé

Cet article propose une méthode de classification sémantique des noms qui repose sur l'application de tests distributionnels. Sont présentées 14 classes sémantiquement simples regroupées en trois pôles : (i) les noms dénotant des entités (comme *cerf*, *pont*, *table*), (ii) les noms dénotant des situations (comme *promenade*, *intelligence*, *délabrement*) et (iii) les noms relevant de classes en marge des deux précédentes (comme *minute*, *hypothèse*, *euro*). S'y ajoutent 7 classes complexes permettant de rendre compte du sens hybride que présentent nombre de noms, tel que, par exemple, le nom *déposition* qui renvoie de manière combinée à une action et à un objet cognitif (*Elle a lu la déposition effectuée par le jeune homme*). Cette classification permet non seulement d'attribuer aux noms, de manière contrôlée, des types sémantiques linguistiquement pertinents, mais également de traiter les différents cas de noms à sens multiples (homonymie, polysémie, hybridité).

Keywords: classification nominale; type sémantique; polysémie; tests distributionnels; coprédication

Introduction

Dans cet article¹, nous présentons une classification sémantique à gros grain des noms du français, fondée sur des tests linguistiques. Les entreprises de classification comportent, comme on le sait, des limites, qui tiennent notamment à la nécessaire discrétisation de phénomènes distribués sur un continuum. La classification des faits en extension peut poser problème, car elle requiert souvent une représentation rigidifiée de certaines propriétés. Les classifications

¹Nous remercions les relecteurs ou relectrices anonymes, les éditeurs de JFLS, ainsi que Justine Salvadori, pour leurs remarques sur une version antérieure de ce travail. Les erreurs restantes nous sont imputables.

présentent néanmoins l'intérêt de proposer une grille d'analyse qui permet d'appréhender les éléments à classer et d'identifier leurs propriétés distinctives. *In fine*, le recours aux catégories, si celles-ci sont envisagées en tant que pôles de classification, apparaît plus éclairant que problématique théoriquement. De fait, les travaux en linguistique font régulièrement appel aux classifications sémantiques, que ce soit à finalité directe ou indirecte (s'agissant de rendre compte de faits connexes, tels que les restrictions de sélection pesant sur les arguments d'un prédicat ou les opérations sémantiques associées à la construction morphologique).

Il existe dans la littérature des propositions de classification sémantique des noms en français (Gross 1996, Flaux et Van de Velde 2000, Huyghe 2015), et certaines classes ont fait l'objet de travaux spécifiques (Balibar-Mrabti 1995, Van de Velde 1995, Godard et Jayez 1996, Grossmann et Tutin 2005, Haas *et al.* 2008, Goossens 2009, Flaux et Stosic 2012, entre autres). L'identification des catégories distinguées n'est cependant pas toujours étayée par le recours à des tests linguistiques systématisés. Par ailleurs, la classification des lexèmes dans les travaux théoriques est rarement envisagée en extension, avec pour objectif une couverture lexicale complète ni basée sur une méthodologie de classification. Inversement, les ressources lexicales ou lexicographiques existantes proposent des descriptions couvrantes, mais les critères de classification sont rarement explicités. La description sémantique des noms repose souvent sur la seule évaluation référentielle, et donne l'impression de s'apparenter à un classement des objets du monde plutôt qu'à l'identification de propriétés linguistiques distinctives (cf. par exemple WordNet pour l'anglais, Petrolito et Bond 2014).

Nous adoptons ici une approche distributionnaliste, en présupposant que les propriétés sémantiques d'une unité lexicale déterminent son environnement linguistique et que celui-ci peut en retour nous renseigner sur le sens des unités lexicales (Harris 1954). Il faut noter toutefois que, considérés individuellement, tous les contextes d'emploi d'un mot ne sont pas également distinctifs et caractéristiques de ses particularités sémantiques. Les tests proposés dans ce travail reposent précisément sur la compatibilité avec des constructions propres à certaines catégories sémantiques. Ils suggèrent une hiérarchisation des contextes d'emploi selon le degré de spécificité sémantique que ceux-ci peuvent révéler. Le recours aux tests distributionnels vise un certain contrôle de la catégorisation, même s'il comporte lui aussi des limites avérées. Les tests impliquent, sinon une certaine normativité, tout au moins une évaluation intuitive des tendances d'emploi des unités testées, dont la fiabilité peut être questionnée. La consultation des corpus ainsi que la co-annotation permettent de limiter ce problème, mais aucun système de tests n'évitera l'existence de cas douteux ou exceptionnels, difficiles à analyser selon les critères proposés. Il reste qu'en dépit de ces imperfections, les tests distributionnels garantissent une assise linguistique à la classification sémantique, et constituent un apport indéniable à l'évaluation purement référentialiste. Il ne s'agit pas ici de disqualifier les conceptions référentielles du sens nominal, mais de les mettre en regard d'une certaine matérialité linguistique, permettant de guider l'analyse sémantique, et

d'éviter que celle-ci ne soit cantonnée à une réflexion sur la nature des référents plutôt que sur celle des mots².

L'objectif spécifique de cet article est d'exposer une typologie sémantique des noms et de proposer une méthode permettant leur annotation sémantique aussi bien en langue qu'en discours. Les tests proposés se fondent sur les propriétés sémantiques identifiées dans la littérature et s'inspirent des tests existants, avec certains amendements, ajouts et propositions nouvelles. Le travail présenté est un état de notre réflexion après plusieurs campagnes d'annotation lexicale réalisées dans le cadre de différents projets et études, en particulier du projet Nomage (Balvet et al. 2011), de l'étude des noms morphologiquement simples (Tribout et al. 2014), et du projet FR-SemCor (Barque et al. 2020).

L'article est organisé comme suit : dans la première section, nous présentons les principes généraux de classification ; dans la deuxième section, nous présentons les 14 classes simples retenues dans notre étude, avec les tests associés ; la troisième section recense les 7 classes complexes qui regroupent des noms sémantiquement hybrides (i.e. relevant conjointement de plusieurs types sémantiques) et les tests de co-prédication associés ; enfin, la quatrième section expose la méthode de classification.

1. Principes de classification

Nous commençons par présenter quelques éléments nécessaires à la bonne compréhension de la méthode proposée. Ceux-ci concernent l'emploi des termes désignant les classes de noms (section 1.1), le fonctionnement général des tests (section 1.2) et le traitement des noms à sens multiples (section 1.3).

1.1. Métalangue et langue ordinaire

Les étiquettes employées pour nommer les classes sont des objets métalinguistiques qui ne présupposent pas de cohérence avec l'emploi standard des termes. Ainsi, on désigne par « noms d'ÉVÉNEMENT », « noms d'ANIMÉ », « noms d'OBJET NATUREL », etc. des noms qui possèdent une série de traits linguistiques qui les fait appartenir aux classes commodément appelées ÉVÉNEMENT, ANIMÉ, OBJET NATUREL. Par exemple, les noms d'ÉVÉNEMENT appartiennent à cette classe en vertu de leurs propriétés sémantiques (ce sont des noms abstraits qui dénotent des situations d'aspect dynamique), repérables par leur comportement distributionnel qui leur permet de valider les tests afférents à cette classe, et non en vertu de l'adéquation référentielle avec ce que nous appelons un "événement" dans la vie quotidienne. Ainsi, selon les tests appliqués, un nom comme *réparation* sera classé parmi les noms d'ÉVÉNEMENT, alors qu'on ne se représente pas nécessairement une réparation comme un "événement" dans le sens courant du terme. Corollairement, les propriétés des substantifs employés métalinguistiquement pour nommer les

²Le recours à des tests distributionnels n'exclut pas que les mots soient dotés de propriétés référentielles. Notre idée est que les critères de construction linguistique sur lesquels notre classification est fondée sont en adéquation avec certaines propriétés sémantiques, et que ces dernières correspondent à la dénotation de différents types de référents. En ce sens, les classes distributionnelles que nous proposons sont associées à diverses classes référentielles, sur la base de propriétés descriptives distinctes.

catégories sémantiques ne nous renseignent pas sur les propriétés des noms qui instancient ces catégories. Afin d'éviter toute ambiguïté entre langue ordinaire et métalangue, les noms choisis pour nommer les classes sémantiques apparaissent dans le texte en petites capitales.

1.2. Fonctionnement des tests

La classification sémantique que nous proposons s'appuie sur les propriétés distributionnelles des noms mises en évidence par les tests, avec l'idée (i) que les différentes constructions sont sémantiquement révélatrices, et (ii) que certaines constructions sont fortement discriminantes³. La personne qui applique un test prend position sur son applicabilité pour une acception donnée d'un nom. Afin de limiter le recours à l'intuition des annotateurs, il est nécessaire de consulter de vastes corpus du français pour déterminer l'acceptabilité des phrases résultant de l'application des tests. Pour réaliser le travail que nous présentons ici, nous avons eu recours aux corpus *frCow* (Schäfer et Bildhauer 2012, Schäfer 2015), *frTenTen12* et *frTenTen17* (Jakubicek et alii 2013) qui ont été interrogés avec les outils d'exploration « SketchEngine » et « NoSketchEngine ». Il s'agit de corpus constitués d'extractions du Web (2016 pour *FrCow*; 2012 et 2017 pour les corpus *frTenTen*) qui offrent l'avantage d'être très vastes (entre 6 et 10 milliards de mots selon le corpus). Notons que le recours aux corpus pour déterminer les acceptabilités est à la fois qualitatif (à travers une évaluation des exemples attestés) et quantitatif (le nombre d'exemples recueillis attestant de la représentativité de la distribution observée).

Les réponses possibles aux tests sont "oui" (o) ou "non" (n). Les tests proposés sont tous complexes, en ce sens qu'ils ne se résument pas à une tournure unique à appliquer au nom. Différentes constructions sont proposées et selon les cas, celles-ci doivent être vérifiées de manière conjointe ou disjointe. Un test conjonctif (marqué "ET") est positif si le N testé peut entrer dans l'ensemble des constructions distributionnelles proposées. La conjonction permet que le test ne soit pas trop permissif et n'accepte que les noms de la classe visée. Un test disjonctif (marqué "OU") est positif si le N testé peut entrer dans au moins l'une des tournures proposées. La disjonction permet d'augmenter la couverture de la classe.

1.3. Traitement des noms à sens multiples

On sait que les noms, notamment les plus fréquents (cf. Zipf 1949), sont susceptibles d'avoir plusieurs sens⁴. À titre indicatif, 74% des 10 000 noms les plus employés dans

³L'idée que des tests distributionnels peuvent permettre d'identifier certains types de noms est défendue notamment par Gross (1994, 2008), dans le modèle des « classes d'objets ». Ces dernières sont des catégories nominales, parfois très précises (ex. FACTURE, COMBUSTIBLE, AÉRONEF, ANIMAL DE BOUCHERIE), définies par la construction avec des prédicats spécifiques dans une interprétation donnée. Élaborées pour permettre d'isoler les différentes significations des verbes avec lesquels elles se combinent, les classes d'objets présentent toutefois un risque de circularité. Si les noms constituent l'entrée sémantique à partir de laquelle sont établies les acceptations des prédicats verbaux, on peut se demander quels éléments linguistiques, autres que le critère référentiel, déterminent en premier ressort la classification des noms.

⁴Pour un état de l'art sur la polysémie et les phénomènes de sens multiples apparentés, voir Kleiber (1999), Falkum et Vicente (2015), entre autres.

frWaC, un large corpus du Web français (1,3 milliard de mots collectés dans le domaine .fr en 2010), reçoivent plusieurs définitions dans le *Wiktionnaire*⁵. Toute entreprise de classification sémantique doit donc prendre en compte ce phénomène massif.

Parmi les noms à sens multiples, il convient de distinguer ceux dont les sens sont exclusifs les uns des autres (en cas d'homonymie ou de polysémie) de ceux qui ont des sens multiples compatibles entre eux, et que nous appellerons ici les noms sémantiquement hybrides⁶. Pour savoir si des sens associés à une même forme sont ou non compatibles, on recourt traditionnellement au test de la co-prédication (Cruse 1986, Godard et Jayez 1996, Asher et Pustejovsky 2006, Asher 2011, Murphy 2021, entre autres). Ce test consiste à appliquer à une même occurrence du nom testé deux prédicats ciblant chacun un sens différent. Si la co-prédication construite produit un effet zeugmatique, comme avec le nom *bureau* en (1c), alors les deux sens testés, celui de meuble en (1a) et celui de pièce de travail en (1b), sont incompatibles. Si au contraire la co-prédication est jugée naturelle, comme avec *livre* en (2c), c'est que les deux sens testés, celui de contenu informationnel en (2a) et de support physique en (2b), sont compatibles entre eux.

- (1) a. Il a acheté un **bureau** en acajou avec plusieurs tiroirs.
 b. Son **bureau** dispose d'une fenêtre.
 c. ? Son **bureau**, qui est en acajou avec plusieurs tiroirs, dispose d'une fenêtre.
- (2) a. J'ai beaucoup lu ce **livre**.
 b. Ce **livre** a été abîmé par l'eau.
 c. Ce **livre**, que j'ai beaucoup lu, a été abîmé par l'eau.

En termes de classification, les noms comme *bureau* relèveront d'autant de classes qu'ils ont de sens distincts et incompatibles entre eux, tandis que les noms hybrides comme *livre* relèveront de classes complexes, dont les plus fréquentes seront présentées en détail dans la section 3.

Soulignons enfin que le traitement des sens multiples dépend de la granularité de la classification. Plus les classes sont générales, plus le nombre de sens distincts par nom considéré diminue. Le nom *domino*, par exemple, peut dénoter une pièce d'un jeu (*le joueur qui a le domino double le plus fort commence*) ou une pièce utilisée en électricité (*changer un domino sur une applique*). Dans une classification purement ontologique et qui n'identifie qu'une seule classe générale pour les noms d'ARTEFACT, cette polysémie du nom *domino* n'apparaîtra pas. La classification que nous présentons ici s'appuie en effet sur des classes générales et laisse donc dans l'ombre des distinctions de sens qui peuvent être plus ou moins importantes, selon la visée de chaque étude. Ce choix est essentiellement guidé par des contraintes d'ordre pratique : l'objectif étant ici de pouvoir appliquer la

⁵Le comptage a été opéré à partir de la version 2015 utilisée dans la ressource *Glawi* (Hathout et Sajous 2016).

⁶Ces noms ont reçu différentes appellations dans la littérature : "noms à facettes" (Cruse 1995), "noms à types pointés" (Pustejovsky 1995), "noms sémantiquement ambivalents" (Milicevic et Polguère, 2010). Ils sont tantôt présentés comme un cas particulier de polysémie, alors dite "logique" ou "inhérente" (Copestake et Briscoe 1995, Frisson 2015), tantôt comme un cas particulier de monosémie impliquant un concept complexe (Kleiber 1999, Cruse 2000).

classification à l'ensemble du lexique nominal du français, nous ne pouvons envisager un grain sémantique plus fin sans considérablement alourdir la description sémantique. Il est néanmoins possible d'appliquer dans un second temps la même méthodologie aux noms des classes générales, une fois identifiés les tests permettant d'en distinguer des sous-classes.

2. Définition des classes simples

Cette section est consacrée à la présentation des 14 classes simples de notre typologie. Les classes considérées sont logiquement complémentaires, c'est-à-dire qu'elles sont appréhendées selon un même principe de catégorisation. En l'occurrence, la classification que nous présentons est de type ontologique, et non relationnel, c'est-à-dire qu'elle repose sur une partition du réel de référence, et non sur la description des relations sémantiques impliquées par les noms. Les catégories envisagées sont par exemple celles de nom d'ANIMÉ, nom d'ARTEFACT ou nom d'ÉVÉNEMENT, et non celles de nom collectif, nom partitif ou nom relationnel, qui sont logiquement compatibles avec les premières, et qui doivent faire l'objet d'une identification et d'une typologie indépendantes (Huyghe 2015).

Nous présenterons tout d'abord les classes de noms ayant en commun de dénoter des entités concrètes pourvues d'une étendue dans l'espace, à savoir les noms d'ANIMÉ, d'ARTEFACT et d'OBJET NATUREL (section 2.1). Viendront ensuite les noms dénotant des situations, parmi lesquelles on distingue les situations dynamiques, regroupant les noms d'ÉVÉNEMENT et ceux de DOMAINE, des situations statives, regroupant les noms de PROPRIÉTÉ et ceux d'ÉTAT (section 2.2). Nous listerons enfin les classes que l'on peut considérer comme secondaires en ce qu'elles constituent des cas intermédiaires vis-à-vis des classes fondamentales et/ou dont l'extension est plus restreinte: noms d'INSTITUTION, de PHÉNOMÈNE, d'OBJET COGNITIF, d'OBJET FINANCIER, de TEMPS, de QUANTITÉ et de MALADIE (section 2.3).

2.1. Les noms dénotant des entités

Linguistiquement, une première ligne de partage se fait entre les N d'entité et les autres types de N selon les critères conjoints de localisation et d'étendue spatiales, ce qui correspond à l'opposition concret *vs* abstrait, dans une conception possible de cette distinction (cf. Kleiber et Vuillaume 2011). En effet, les noms se répartissent entre ceux qui peuvent dénoter des entités localisables dans l'espace, ce qui se manifeste par la compatibilité avec les compléments de dimension et de mesure (cf. tests T1a et T1b), et ceux qui ne le peuvent pas. Les noms d'entité dénotent toutes sortes d'objets et d'êtres vivants et l'on trouve des tests linguistiques pour les discriminer dans Dowty (1991), Godard et Jayez (1996), Flaux et Van de Velde (2000), Huyghe (2009), Kleiber *et al.* (2012), entre autres.

Afin de discriminer les noms d'entité des autres types de noms, il convient d'appliquer en premier lieu le test conjonctif T1 présenté ci-dessous. Il faut d'une part que le N entre dans la tournure T1a et, d'autre part, qu'il entre dans celle de T1b. On note que pour T1b, trois formulations sont proposées (b1, b2 ou b3). Elles ne correspondent pas à un choix ni à une disjonction, mais

répondent à des besoins de formulation différents selon que le N à classer est comptable [T1-b1], massif [T1-b2] ou collectif [T1-b3] — concernant les N collectifs (Ncoll), nous retenons la définition qu'en donne Lammert : est dit collectif « un nom qui, au singulier, dénote une entité composée d'un regroupement d'éléments » (Lammert 2010 : 66–70).

T1⁷

a. *Dét N se trouve* + complément de localisation spatiale

ET

b1. *Dét N de* + mesure [pour les N comptables]

b2. mesure + *de N* [pour les N massifs]

b3. *Dét N de Dét.card N₂* [pour les N collectifs]

Dans T1 comme dans l'ensemble des autres tests proposés, sauf mention contraire, l'étiquette *Dét* renvoie à un déterminant singulier d'interprétation spécifique tiré de la liste suivante : *ce / cet / cette, le / la, un / une, du / de la. Dét.card* est à un déterminant cardinal nécessairement pluriel (*deux, trois, quinze, deux cents, etc.*). En T1a, on n'acceptera pas n'importe quel complément de localisation, mais uniquement ceux qui dénotent des lieux physiques, et que l'on piochera de préférence dans la liste (3). En T1b, l'élément "mesure" correspond à une forme composée d'un déterminant cardinal et d'un nom dénotant une unité de mesure. On s'inspirera des exemples proposés en (4a) pour les N comptables, en (4b) pour les N massifs et en (4c) pour les N collectifs.

- (3) sur la table, à côté du sac, dans le jardin, à Paris, entre la table et le mur, en Europe, près du poumon droit, *etc.*
- (4) a. un N de {deux mètres de large / vingt mètres carrés / trois mètres cubes / cent hectares / cinq cents grammes / dix kilos / trente tonnes}
- b. {deux mètres / vingt mètres carrés / trois mètres cubes / cent hectares / cinq cents grammes / dix kilos / trente tonnes} de N
- c. un N de quinze {chèvres / loups / personnes / timbres / pierres / tout N approprié au Ncoll⁸}

Des noms comme *table* (5), *farine* (6) ou *archipel* (7) valident bien le test conjonctif T1a et T1b. Tel n'est pas le cas en revanche du nom *idée* car d'une part, le complément de localisation dans (8a) ne dénote pas un lieu physique, et d'autre part la seconde partie du test n'est pas validée (8b).

- (5) a. Cette **table** se trouve dans le jardin. [T1a]
b. une **table** de deux mètres de large [T1-b1]

⁷Lorsqu'un test est conjonctif, comme T1, les exemples illustratifs sont présentés en batterie (1a), (1b), *etc.* Cette présentation symbolise le fait que l'ensemble des tournures doivent être validées conjointement par le nom pour entrer dans la classe. Les exemples (5) à (8) illustrent cette notation pour T1a et b.

⁸Certains N collectifs spécifient le type d'éléments rassemblés et ont en conséquence un ou plusieurs "noms appropriés" avec lesquels ils forment une collocation naturelle. Le N approprié réfère aux éléments constitutifs prototypiques de la collection dénotée, par exemple *île, abeille, fleur* pour, respectivement, *archipel, essaim, bouquet* (cf. Lammert et Lecolle 2014 : 4-5).

- (6) a. La **farine** se trouve sur la table. [T1a]
 b. deux kilos de **farine** [T1-b2]
- (7) a. L'**archipel** se trouve en Europe. [T1a]
 b. un **archipel** de quinze îles [T1-b3]
- (8) a. Cette **idée** se trouve à la fin du chapitre.
 b. ? une **idée** de dix kilos / ? une idée de trois m² [rejet de T1b]⁹

Les noms qui valident T1 dénotent des entités concrètes et se partagent entre trois classes : (i) ANIMÉ (ex. *cerf*), (ii) ARTEFACT (ex. *table*) et (iii) OBJET-NATUREL (ex. *lac*), que nous présentons dans les trois sous-sections ci-après.

2.1.1. Les noms d'ANIMÉ

Les noms d'ANIMÉ dénotent des vivants doués d'intentionnalité. Les vivants non doués d'intentionnalité (ex. *bactérie*, *virus*, *fleur*) sont dénotés par des noms d'OBJET-NATUREL (cf. 2.1.3). Les noms d'ANIMÉ se distinguent des autres N concrets par la validation de T2. Ils décrivent des entités dotées d'une étendue spatiale et localisables dans l'espace physique [T1a et b], volitionnelles [T2a et b] et capables de s'alimenter [T2c].

T2

a. Dét N {décide / choisit} de SV

ET

b. Dét N + SV {volontairement / délibérément / consciemment}

ET

c. Dét N a beaucoup mangé

Le nom *cerf* valide bien chaque élément du test, comme le montre (9).

- (9) a. Le **cerf** décide de sauter dans l'eau. [T2a]
 b. Le **cerf** a volontairement sauté dans l'eau. [T2b]
 c. Le **cerf** a beaucoup mangé. [T2c]

2.1.2. Les noms d'ARTEFACT

Parmi les noms qui sont positifs à T1 et négatifs à T2, c'est-à-dire parmi les noms qui dénotent des choses non animées, il est linguistiquement pertinent de distinguer ceux qui décrivent des artefacts de ceux qui décrivent des objets naturels. Parmi les premiers, on peut lister les noms d'objet, comme *table* ou *robe* (10), les noms de document, comme *page de garde* (10), les noms de construction, comme *pont*

⁹Le symbole « ? » signale que le test n'est pas validé. Le choix de ce symbole, moins contraignant que l'astérisque d'agrammaticalité, rappelle que l'acceptabilité est une notion relative et que la coercition du sens lexical en contexte reste souvent possible. Ce qui est toutefois visé ici, c'est la (non-)canonicité des énoncés testés. L'existence de productions isolées (en nombre) et marginales (i.e. sémantiquement incongrues) ne devrait pas suffire pour considérer qu'un test est véritablement validé.

(11) ou encore les noms de mets, comme *gâteau* et *glace* (12). Tous ces noms valident le test disjonctif T3 présenté ci-dessous.

T3¹⁰

a. $N_{[humain]}$ a {fabriqué / déchiré} Dét N {à tel endroit / à tel moment}

OU

b. $N_{[humain]}$ a construit Dét N + Compl de matière + Adj de couleur

OU

c. $N_{[humain]}$ a confectionné Dét N + Compl d'arôme

En T3b, le complément de matière correspond à un syntagme prépositionnel introduit par la préposition *en* suivie d'un nom dénotant une matière ou un matériau (ex. *en bois*, *en plastique*). En T3c, le complément d'arôme correspond à un syntagme prépositionnel introduit par la préposition *à* suivie d'un article défini et d'un nom dénotant un arôme ou un parfum (ex. *à la vanille*, *au rhum*).

- (10) Pierre a fabriqué cette table l'année dernière. / Pierre a déchiré {la **page de garde** / la **robe** de sa sœur} hier. [T3a]
 (11) Jeannette a construit ce **pont** en pierre grise au-dessus de la rivière. [T3b]
 (12) Jo a confectionné {un **gâteau** au chocolat / une **glace** à la rose / de la **purée** au cumin}. [T3c]

2.1.3. Les noms d'OBJET-NATUREL

Parmi les N concrets, certains dénotent des objets naturels qui ne sont ni des artefacts (puisque non fabriqués par l'homme) ni des animés (puisque non doués d'intentionnalité). Ces noms d'OBJET-NATUREL (OBJ-NAT) décrivent des substances naturelles (ex. *neige*, *sang*, *graphite*), des organismes vivants sans intentionnalité (ex. *plante*, *fleur*, *bactérie*), ou des lieux naturels (ex. *mont*, *plaine*, *lac*, *glacier*). Ces noms valident T1 (13a-b), mais rejettent T2 (13c-d) et T3 (13e-f).

- (13) a. Le **mont** se trouve près de la frontière algérienne. [T1a]
 b. un **mont** de 200 mètres de haut [T1b]
 c. ? Le **mont** a décidé de passer la frontière. [rejet de T2a]
 d. ? Le **mont** a volontairement passé la frontière. [rejet de T2b]
 e. ? Jean a fabriqué un **mont** dans son jardin. [rejet de T3a]
 f. ? Julie a construit un **mont** en plastique bleu. [rejet de T3b]

¹⁰Lorsqu'un test est disjonctif, comme T3, les exemples illustratifs sont présentés en série, chacun portant un numéro distinct : (1), (2), etc. Cette présentation symbolise le fait que la validation d'une seule des tournures suffit pour que le nom entre dans la classe.

2.2. Les noms dénotant des situations

Il existe plusieurs classes de noms renvoyant à des situations¹¹, qui peuvent être d'aspect dynamique ou statif, et que nous présentons dans cette section¹². Pour distinguer les classes en question, nous utilisons et aménageons certains tests classiques proposés dans les travaux existants (cf. Gross 1981, Gross et Vivès 1986, Balibar-Mrabti 1990, Gross et Kiefer 1995, Van de Velde 1995, Godard et Jayez 1996, Kiefer 1998, Daladier 1999, Gaatone 2004, Tutin *et alii* 2006, Beausery 2009, Haas 2009, Huyghe 2014, entre autres).

2.2.1. Les noms d'ÉVÈNEMENT

Une classe importante est celle qui regroupe les noms dénotant des situations dynamiques, c'est-à-dire des situations dans lesquelles il se passe quelque chose, soit qu'un événement survient, soit qu'une action est réalisée. Ces noms relèvent tous de la classe ÉVÈNEMENT, étiquette métalinguistique qui ne s'oppose pas ici à "action", mais qui l'englobe. Les noms entrant dans cette classe valident le test T4 présenté ci-dessous.

T4

a. *Dét N {a eu lieu / s'est produit} {à tel moment / à tel endroit}*

OU

b. *N_[humain] {a procédé à / a accompli / a effectué} Dét N (+ expansion facultative)*

(14) Le **séisme** a eu lieu hier à Toronto. [T4a]

(15) Le mécanicien a accompli une [T4b]
réparation délicate avec succès.

2.2.2. Les noms de DOMAINE

Les noms de DOMAINE regroupent des noms qui dénotent des activités (*jardinage*) et des champs de connaissance (*histoire*), en vertu du fait que ces deux types de dénotation sont souvent inextricablement liés. On associe un domaine d'expertise aux activités, et des activités (physiques ou intellectuelles) à de nombreux domaines d'expertise. A *minima*, les noms de DOMAINE vérifient T5a ou T5b, mais de nombreux noms peuvent valider ces deux tests conjointement.

¹¹Les noms de situation sont des noms abstraits pourvus d'aspect lexical. Il s'agit pour l'essentiel de noms "syncatégorématiques", c'est-à-dire des noms qui « ne peuvent pas, par eux-mêmes, [...] repérer les référents particuliers auxquels ils peuvent s'appliquer » (Kleiber, 1981 : 40). À l'inverse, les noms rassemblés dans la section 2.1. (ANIMÉ, ARTEFACT, OBJET-NATUREL) sont des noms « catégorématiques ».

¹²La distinction dynamique / statif est une distinction aspectuelle d'abord étudiée dans le domaine verbal. De nombreuses classifications aspectuelles des prédicats verbaux sont proposées dans la littérature, pour un aperçu, on pourra utilement consulter le tableau récapitulatif proposé par Asnès et Kupferman (2008 : 28).

T5

a. $N_{[humain]}$ fait {du / de la} N

OU

b. $N_{[humain]}$ a reçu une formation dans le domaine {du / de la} N

(16) Louise fait {du **jardinage** / du **journalisme** / de la **natation**}. [T5a]

(17) Louise a reçu une formation dans le domaine {du **journalisme** / de la **philosophie** / de l'**ophtalmologie**}. [T5b]

Précisons que le verbe *faire* utilisé en T5a ne doit avoir ni le sens de ‘fabriquer/confectionner’, ni celui de ‘avoir’. Ainsi, les noms *purée* (18) ou *eczéma* (19) ne sont pas des noms de DOMAINE, mais respectivement un nom d’ARTEFACT (cf. *supra* section 2.1.2) et un nom de MALADIE (cf. *infra* section 2.3.7)

(18) René fait de la **purée** de carotte. = René a confectionné de la purée de carotte.

(19) Pierre fait de l'**eczéma**. = Pierre a de l'eczéma.

Si nous présentons les noms de DOMAINE dans la partie consacrée aux situations, c’est d’une part parce que ces noms rejettent les tests propres aux N concrets (T1 à T3), tout comme ceux spécifiques des classes secondaires (cf. *infra* section 2.3), et d’autre part parce qu’une partie des noms de cette classe sont aspectuellement analysables comme dynamiques. Les noms tels que *jardinage*, *natation*, *muscultation*, *patinage* peuvent être analysés comme relevant de la classe aspectuelle des « activités¹³ » (cf. Van de Velde 1995, Haas et Huyghe 2010). Ils sont dynamiques en vertu même de leur compatibilité avec T5a, *faire du* étant connue comme structure permettant de mettre en évidence le trait de dynamicit  dans le domaine nominal (cf. Giry-Schneider 1978 et 1987, Van de Velde 1997, Vigier 2003, Heyd et Knittel 2009). Ces noms  tant fondamentalement massifs, ils valident T5a (20), mais ne v rifient pas les tests propos s pour la classe  V NEMENT (21–22), qui conviennent strictement aux noms comptables (Huyghe 2011).

(20) Louise a fait du **jardinage** hier matin. [T5a]

(21) ? Le **jardinage** a eu lieu ce matin. [rejet de T4a]

(22) ? Pierre a {accompli / proc d   } une **natation** efficace. [rejet de T4b]

Cela justifie l’existence de la classe des noms de DOMAINE, distincte de celle des  V NEMENT mais incluse dans les noms d notant des situations.

¹³Le terme *activit * est emprunt  aux classifications d’Aktionsart verbale (*activity* en anglais, cf. Garey 1957, Vendler 1967). Il d note un type aspectuel caract ris  par la dynamicit , la dur e et l’at licit  des situations d crites.

2.2.3. *Les noms de PROPRIÉTÉ*

Parmi les situations statives, qui rejettent à la fois les tests de dynamicité T4 et T5a, on distingue la classe des N de PROPRIÉTÉ et celle des N d'ÉTAT (cf. *infra*, section 2.2.4). Cette distinction s'apparente à l'opposition entre "individual-level predicate" (ILP) et "stage-level predicate" (SLP) (Carlson 1977) qui sépare les propriétés structurelles, considérées comme inhérentes et non soumises au temps, des propriétés transitoires, considérées comme contingentes et extrinsèques¹⁴. Les noms de PROPRIÉTÉ, qui correspondent aux ILP et expriment des propriétés essentielles, sont identifiés par le test disjonctif T6 présenté ci-dessous.

T6

a. N_0 est d'un(e) grand(e) N

OU

b. $N_{[humain]}$ a fait preuve {de N / d'un(e) N + expansion}

OU

c. Cela m'a altéré Dét N

(23) Jeanne est d'une grande **intelligence**. [T6a](24) Cette maison est d'une grande **insalubrité**. [T6a](25) Pierre a fait preuve de **couardise**. [T6b](26) Cela m'a altéré {le **goût** / la **mémoire**}. [T6c]

Notons que T6a, si le sujet est un animé, doit être paraphrasable par « avoir du N », comme c'est le cas dans (27), ce qui permet d'exclure des noms comme *famille* ou *aide*. Ces derniers peuvent en effet être insérés dans la tournure T6a, mais les phrases résultantes ne sont pas synonymes de la construction en *avoir* (28).

(27) a. Jeanne est d'une grande **intelligence**. = Jeanne a de l'**intelligence**. [T6a](28) a. Pierre est d'une grande **famille**. ≠ Pierre a de la **famille**.b. Pierre est d'une grande **aide** pour s'occuper des enfants.≠ Pierre a de l'**aide** pour s'occuper des enfants2.2.4. *Les noms d'ÉTAT*

Les noms d'ÉTAT correspondent aux SLP, et vérifient le test disjonctif T7. Ces noms décrivent des situations transitoires (psychologiques ou matérielles) dans lesquelles peuvent se trouver des entités animées (29) ou non animées (30).

T7

a. $N_{[humain]}$ {ressent / éprouve} Dét N (+ expansion)

OU

b. $N_{[inanimé]}$ est dans un état de N (+ expansion)

¹⁴Cette opposition est proposée pour le français avec d'autres terminologies, par exemple, « propriétés intrinsèques » vs « propriétés extrinsèques » (Anscombe, 1995 : 45), « qualités » vs « états » (Van de Velde, 1995 : 129-220).

(29) Pierre {ressent / éprouve} {de la **crainte** / une vive **crainte**}. [T7a]

(30) La maison est dans un état de **délabrement** avancé. [T7b]

2.3. Les classes secondaires

Dans cette section, sont regroupées 7 classes de noms dont le comportement linguistique est original et suffisamment distinct de celui des classes principales (noms d'entité et noms de situation) pour constituer des classes sémantiques autonomes. Certaines de ces classes ont également une extension plus restreinte que celles dénotant des entités ou des situations. Ces classes, bien que relevant d'un grain sémantique plus fin que les classes précédentes, n'entretiennent pas avec elles de rapport hiérarchique inclusif et n'en constituent donc pas des sous-classes.

2.3.1. Les noms d'INSTITUTION

Les noms d'INSTITUTION, qui ont fait l'objet d'une attention particulière dans le cadre des études portant sur les entités nommées (Ehrmann 2008, Grouin *et al.* 2011, entre autres), sont des noms dénotant des organismes, des commerces, des organisations politiques, administratives, scolaires, financières ou associatives. La validation du test disjonctif T8 permet de les identifier.

T8

a. *Dét N a été fondé en telle année*

OU

b. *N_[humain] a été nommé à la tête de (Dét) N*

(31) {Cette **banque** / L'**académie** / Ce **restaurant**} a été fondé(e) en 1800. [T8a]

(32) Pierre a été nommé à la tête {de l'**ONU** / de la **faculté** / du **conseil**}. [T8b]

Les noms d'INSTITUTION ont en commun avec les noms d'ANIMÉ de valider les tests d'intentionnalité T2a et T2b (33). En revanche, ils ne valident pas T2c qui renvoie à la capacité de manducation propre aux animés (34).

(33) a. La **banque** a décidé d'accorder un crédit exceptionnel aux entreprises. [T2a]

b. La **banque** a volontairement diminué les taux d'emprunt. [T2b]

(34) ? La **banque** a beaucoup mangé. [rejet T2c]

2.3.2. Les noms de PHÉNOMÈNE

Les noms de PHÉNOMÈNE décrivent des référents accessibles aux sens, mais qui ne semblent se réduire ni à des objets ni à des événements. Ils dénotent plus précisément des odeurs (ex. *fragrance*), des sons (ex. *cliquetis*) ou des phénomènes lumineux (ex. *leur*), et valident l'une des paires de tests présentées en T9a, T9b ou T9c.

T9
a1. <i>On a entendu Dét N (tonitruant / strident / mélodieux / aigu / grave).</i>
ET
a2. <i>Dét N (a retenti / a résonné / s'est propagé) (au fond du couloir / dans la cour).</i>
OU
b1. <i>On a vu Dét N (aveuglant / éblouissant / blafard / pâle).</i>
ET
b2. <i>Dét N (a resplendi / s'est propagé / a scintillé / a jailli) au fond du couloir.</i>
OU
c1. <i>On a senti Dét N (nauséabond / âcre / enivrant / tenace / entêtant).</i>
ET
c2. <i>Dét N (s'est répandu / persiste / embaume) (dans) la pièce.</i>

Les noms de PHÉNOMÈNE ne dénotent ni des entités matérielles, comme le montre le rejet de T1 en (35a), ni des situations dynamiques, comme le montre le rejet de T4 en (35b). Les phénomènes sonores valident T9a (36), les phénomènes lumineux valident T9b (37) et les phénomènes olfactifs valident T9c (38)¹⁵.

- (35) a. ? {Le **son** / la **lumière** / l'**odeur**} se trouve au fond du couloir [rejet de T1]
 b. ? {Un **son** / une **lumière** / une **odeur**} {s'est produit(e) / a eu lieu} au fond du couloir. [rejet de T4]
- (36) a. On a entendu un **son** aigu. [T9-a1]
 b. Un **son** a retenti au fond du couloir. [T9-a2]
- (37) a. On a vu une **lumière** aveuglante. [T9-b1]
 b. Une **lumière** a jailli au fond du couloir. [T9-b2]
- (38) a. On a senti une **odeur** nauséabonde. [T9-c1]
 b. Cette **odeur** persiste dans la pièce. [T9-c2]

2.3.3. Les noms d'OBJET-COGNITIF

Les N d'OBJET-COGNITIF (OBJ-COG) dénotent des contenus informationnels (ex. *idée, hypothèse*). Ils sont aussi appelés "objets informationnels" (Godard et Jayez 1996), et correspondent à une partie de ce que Flaux et Stosic (2014) appellent des "noms d'idéalités concrètes". Les trois tournures proposées en T10 permettent, puisque leur emploi est disjonctif, de rassembler au sein de la classe des N d'OBJET-COGNITIF des noms couvrant un spectre sémantique varié, incluant notamment des noms en lien avec des V d'action (*affirmer – affirmation, délirer – délire*), des objets textuels (*poème*), culturels et artistiques (*pièce de théâtre, chanson*), des noms exprimant un jugement (*évidence*), etc.

¹⁵Concernant les odeurs, voir par exemple Kleiber et Vuillaume (2011), Kleiber (2012) et, pour une approche cognitive, David (2000 : 76-77) qui classe les odeurs parmi les phénomènes ou les substances sur la base de la représentation qu'en ont les locuteurs ; à propos du caractère intensif des « sensations physiques » que sont le bruit, la lumière et les odeurs, cf. Kleiber (2013 : 72 et ss).

T10

a. *Dét N selon lequel P est difficile à admettre*

OU

b. *N_[humain] écoute Dét_[plur] N décousus de Vincent*

OU

c. *N_[humain] a écrit Dét N {intéressant / pertinent}*

- (39) {L'**évidence** / L'**hypothèse** / L'**idée**} selon laquelle l'eau manquera demain sur Terre est difficile à admettre. [T10a]
 (40) Marie écoute les {**élucubrations** / **mensonges** / **propos**} décousu(e)s de Vincent. [T10b]
 (41) Pierre a écrit un(e) {**épopée** / **paragraphe** / **sarabande**} intéressant(e). [T10c]

Les N d'OBJET-COGNITIF ont un comportement particulier du point de vue de la localisation spatiale, ce qui est visible dans les exemples proposés par Godard et Jayez (1996). Une phrase comme *Cette idée se trouve déjà dans le Gorgias* (Godard et Jayez 1996: 44) indique certes la localisation d'un objet informationnel (*idée*), mais à l'aide d'un repère qui est lui-même un objet cognitif (*le Gorgias*), ce qui pose un problème de circularité. Le test de localisation T1 (cf. *supra*, section 2.1) évite cet écueil puisqu'il est très contraint et oblige à ce que le complément de localisation soit clairement physique.

2.3.4. Les noms d'OBJET-FINANCIER

Les noms d'OBJET-FINANCIER (OBJ-FIN) regroupent des noms comme les noms de monnaie (ex. *euro*, *dollar*), mais aussi les noms dénotant des valeurs financières ou des sommes d'argent (ex. *tarif*, *salaire*). Les noms de cette classe valident le test disjonctif T11 (42)–(44).

T11

a. *Il a versé Dét N en euros*

OU

b. *Il a obtenu Dét_[indéf, sing] N modique*

OU

c1. *Dét N est la monnaie de tel pays*

ET

c2. *Quel est le taux de change de Dét N ?*

- (42) Il a versé une **mensualité** en euros. [T11a]
 (43) Il a obtenu un **tarif** modique. [T11b]
 (44) a. Le dollar est la **monnaie** des USA. [T11-c1]
 b. Quel est le taux de change du **dollar** ? [T11-c2]

Les noms d'OBJET-FINANCIER ont en commun de référer à une valeur (soit ils sont porteurs d'une valeur, comme *versement* ou *euro*, soit ils fixent une valeur, comme *tarif*

ou *taux*). Les noms de monnaie, inclus dans la classe des objets financiers en raison de leur sème 'valeur' sont à la fois des artefacts par métonymie (*j'ai un euro dans ma poche*) et des objets virtuels dotés d'une valeur (*l'euro a encore perdu de sa valeur*)¹⁶. Ils forment une classe restreinte, mais spécifique, qui valide la paire de tests T11c1-c2.

2.3.5. Les noms de TEMPS

Les noms entrant dans cette classe ne sont pas tous ceux qui ont pu être repérés comme « noms de temps » dans la littérature (Berthonneau 1989), mais uniquement ceux dénotant des segments temporels (ponctuels ou duratifs), identifiables grâce au test T12. La description de l'étendue temporelle constitue dans le sémantisme de ces noms un trait central, voire unique, ce qui les distingue notamment des noms qui dénotent des événements duratifs (ex. *promenade, réunion, concert*).

T12

a. *Ils se sont rencontrés* {N / courant N / Dét N / Dét.card N durant / à N}

OU

b. *Attendez* Dét_[indéf, sing] bref N

- (45) *Ils se sont rencontrés* {**lundi** / courant **janvier** / cet **été** / trois **jours** durant / à **midi**}. [T12a]
 (46) *Attendez un bref* {**instant** / **moment**}. [T12b]

2.3.6. Les noms de QUANTITÉ

Les N de quantité dénotent des unités de mesure (ex. *kilo*), qui valident la paire de tests T13-a1 et T13-a2 (47). La classe contient également d'autres noms tels que les noms dénotant des quantités contenues dans des récipients (ex. *bol (de riz), assiette (de soupe)*), certains événements météorologiques employés métaphoriquement (ex. *avalanche (de mails)*), ou encore des noms collectifs (ex. *forêt (de symboles)*), cf. Benninger 1999), qui ont tous en commun de permettre une quantification et qui valident T13-b (48).

T13

a1. *Dét* {courant / poids / lumière / intensité / tension / distance / période / vitesse} de Dét.card N, *c'est* {peu / beaucoup}

ET

a2. *Dét* N₀ {mesure / fait / pèse / dure} Dét.card N

OU

b. *Dét* N de N₂, *c'est* (peu / beaucoup)

où N₂ appartient à la liste : *farine, huile, sable, messages, images, cheminées, projectiles, juristes*

¹⁶On trouvera quelques remarques sur le rapport complexe entre la valeur de la monnaie et la monnaie en tant qu'objet matériel au détour d'un article de Rastier (2002) traitant centralement de la notion de valeur chez Saussure.

- (47) a. Une distance de 700 **mètres**, c'est beaucoup. [T13-a1]
 b. Cette antenne {mesure / fait} dix **mètres**. [T13-a2]
- (48) a. une **forêt** de cheminées [T13-b]
 b. un **bataillon** de juristes
 c. une **cuillerée** d'huile
 d. un **verre** de farine
 e. une **pelletée** de sable
 f. un **multitude** d'images
 g. une **kyrielle** de messages

2.3.7. Les noms de MALADIE

Les noms de MALADIE dénotent les maux que peuvent subir les animés (Labelle 1986). Ces noms ont un comportement linguistique singulier qui se manifeste dans leurs propriétés distributionnelles, et qui peuvent servir de test T14.

T14

a. *Dét N {est / n'est pas} une maladie contagieuse*

ET

b. *N_[humain] a guéri de Dét N*

ET

c. *Le médecin m'a diagnostiqué Dét_[indef, sing] N*

- (49) a. {Le **tétanos** / L'**eczéma** / L'**appendicite**} n'est pas une maladie contagieuse. [T14a]
 b. Pierre a guéri {du **tétanos** / de l'**eczéma** / de l'**appendicite**}. [T14b]
 c. Le médecin m'a diagnostiqué {un **tétanos** / un **eczéma** / une **appendicite**}. [T14c]

La conjonction des trois structures en T14 permet d'éviter l'extension de la classe des N de MALADIE que pourrait causer l'interprétation métaphorique de la tournure T14a, telle qu'en (50a). Le nom *honte*, qui valide apparemment T14a, mais en aucun cas T14b et T14c, ne sera pas classé parmi les noms de MALADIE, mais parmi les noms d'ÉTAT (*ressentir de la honte*).

- (50) a. La **honte** n'est pas une maladie contagieuse. [T14a]
 b. ? Paul a guéri d'une **honte**. [rejet de T14b]
 c. ? Le médecin m'a diagnostiqué une **honte**. [rejet de T14c]

2.5. Récapitulatif pour l'identification des 14 classes simples

Dans le tableau 1 est récapitulé l'ensemble des 14 classes sémantiques simples présentées ci-dessus et des 14 tests permettant de les identifier. Pour des raisons de présentation, les

Tableau 1 : Classes sémantiques simples et tests identificatoires

Classes \ Tests	Tests													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ANIMÉ	o	o	n	n	n	n	n	n	n	n	n	n	n	n
ARTEFACT	o	n	o	n	n	n	n	n	n	n	n	n	n	n
OBJ-NAT	o	n	n	n	n	n	n	n	n	n	n	n	n	n
ÉVÉNEMENT	n	n	n	o	n	n	n	n	n	n	n	n	n	n
DOMAINE	n	n	n	n	o	n	n	n	n	n	n	n	n	n
PROPRIÉTÉ	n	n	n	n	n	o	n	n	n	n	n	n	n	n
ÉTAT	n	n	n	n	n	n	o	n	n	n	n	n	n	n
INSTITUTION	n	n	n	n	n	n	n	o	n	n	n	n	n	n
PHÉNOMÈNE	n	n	n	n	n	n	n	n	o	n	n	n	n	n
OBJ-COG	n	n	n	n	n	n	n	n	n	o	n	n	n	n
OBJ-FIN	n	n	n	n	n	n	n	n	n	n	o	n	n	n
TEMPS	n	n	n	n	n	n	n	n	n	n	n	o	n	n
QUANTITÉ	n	n	n	n	n	n	n	n	n	n	n	n	o	n
MALADIE	n	n	n	n	n	n	n	n	n	n	n	n	n	o

tests ne sont évoqués que par leur numéro — à toutes fins pratiques, l'ensemble des tests est regroupé en annexe sous forme d'une liste. Nous rappelons que “valider un test” signifie bien entrer dans l'une des tournures proposées pour ledit test (test disjonctif) ou entrer dans l'ensemble des tournures requises (test conjonctif). Pour déterminer la classe simple d'un N, il convient que celui-ci valide le (ou les) tests propres à cette classe (marqués “o”), mais aussi qu'il rejette les tests marqués “n”. Si un N se comporte autrement, alors soit il est polysémique (et il convient de faire passer les tests aux différents sens discriminés en amont), soit il relève d'une classe complexe (cf. section suivante).

3. Définition des classes complexes

Nous présentons dans cette section les sept classes complexes observées pour les noms hybrides du français¹⁷. Pour rappel, ces classes regroupent des noms qui ont la propriété d'avoir des sens distincts, mais compatibles (voir *supra*, section 1.3). Le nom *télégramme*, par exemple, relève de la classe complexe ARTEFACT*OBJET-COGNITIF, puisque ses deux sens de support physique (ARTEFACT) et de contenu

¹⁷Cet inventaire des classes complexes est essentiellement issu des cas décrits dans la littérature et des données du projet FrSemCor (Barque *et al.* 2020). N'ayant pas établi de classification sémantique de l'ensemble des noms du français, nous ne prétendons pas à l'exhaustivité des classes complexes présentées ici.

(OBJET-COGNITIF) sont compatibles entre eux, comme l'indique une co-prédication telle que (51).

(51) Il jeta le **télégramme** qui annonçait la venue du diplomate.

Notons que les tests proposés pour identifier les noms relevant d'une classe complexe sont plus souples que ceux permettant d'évaluer l'appartenance du nom testé à l'une et à l'autre des classes simples formant la classe complexe. On sait en effet que les propriétés de la co-prédication, notamment l'ordre et la nature des prédicats convoqués d'une part, et la nature et la relation entre les deux sens en jeu d'autre part, ont des effets sur le degré d'acceptabilité des phrases construites (Kleiber 1996, Cruse 2004, Gotham 2017, Asher 2011, Ortega-Andrés et Vicente 2019, Murphy 2021, entre autres). Ces variations d'acceptabilité peuvent conduire, pour un même nom, à des analyses contradictoires selon la forme de co-prédication utilisée. Il suffira donc ici que le nom testé puisse entrer dans une occurrence au moins de co-prédication mettant en jeu un prédicat typique de la première classe et un prédicat typique de la seconde classe pour être analysé comme relevant de la classe complexe.

3.1. Les noms hybrides de la classe ARTEFACT*OBJET-COGNITIF

Les noms de cette classe, qui dénotent conjointement un contenu informationnel et son support physique, ont été largement décrits dans la littérature sur les types complexes (Cruse 1995 et 2000, Ortega-Andrés et Vicente 2019, entre autres). Comme le nom *livre*, emblématique de cette classe, des noms comme *tract*, *manuscrit* et *rapport* peuvent entrer dans une co-prédication impliquant un prédicat typique de la facette ARTEFACT et un prédicat typique de la facette OBJET-COGNITIF (52).

(52) a. Dét N [qui se trouve sur la table]_{Pred_ARTEFACT} [a été {écrit / rédigé / lu} par ma sœur]_{Pred_OBJ-COG}.
 b. Le {tract / manuscrit / rapport} qui se trouve sur la table a été rédigé par ma sœur.

3.2. Les noms hybrides de la classe ARTEFACT*INSTITUTION

Les noms dénotant des institutions renvoient à des concepts plus ou moins complexes, qui incluent le plus souvent le bâtiment qui abrite l'institution (53a) et les différents groupes d'individus qui œuvrent en son sein (53b), parmi d'autres sens possibles selon le type d'institution en jeu (Arapinis et Vieu 2015, Ortega-Andrés et Vicente 2019). Le sens d'INSTITUTION est compatible avec celui d'ARTEFACT, comme l'indique la co-prédication en (54). La compatibilité avec le sens d'animé collectif nous semblant plus sujette à caution (55), nous ne postulons pas ici de classe complexe ANIMÉ*INSTITUTION.

- (53) a. Les fenêtres de {la **banque** / l'**hôpital**} ont été cassées.
 b. {La **banque** / L'**hôpital**} a organisé une fête pour son départ.
- (54) a. Dét N [qui se trouve dans cette rue]_{Pred_ARTEFACT} [a été fondé en 1879]_{Pred_INSTITUTION}.
 b. {La **banque** / L'**hôpital**} qui se trouve dans cette rue a été fondée en 1879.
- (55) ? {La **banque** / L'**hôpital**} qui déjeune dans ce restaurant tous les midis a été fondé(e) en 1879.

3.3. Les noms hybrides de la classe ÉVÉNEMENT*ÉTAT

Les noms de cette classe complexe dénotent un type d'action incluant une phase dynamique suivie d'un état résultant, ce qui correspond, d'un point de vue aspectuel, à des achèvements ou des accomplissements "gauches" (Piñón 1997, Apothélos 2008, Fradin 2011, Haas et Jugnet 2013). Dans le test de co-prédication en (56a), la première prédication pointe la facette ÉVÉNEMENT tandis que la facette ÉTAT est identifiée par un prédicat de durée qui doit porter sur l'état résultant, et non sur l'événement (Piñón 1999). Dans le cas de *disparition* en (56b), c'est bien l'absence de la personne qui dure deux semaines et non le passage de l'état présent à l'état absent. Il en va de même pour *emprisonnement* : les deux semaines ne s'appliquent pas à la durée de l'action d'emprisonner, mais bien à la durée de la détention.

- (56) a. Dét.poss N, [qui a eu lieu à tel moment]_{Pred_EVENEMENT} [a duré x temps]_{Pred_ETAT}.
 b. {Sa / son} {**disparition** / **emprisonnement**}, qui a eu lieu cet hiver, a duré deux semaines.

3.4. Les noms hybrides de la classe ÉVÉNEMENT*OBJET-COGNITIF

Les noms de cette classe dénotent conjointement un acte, souvent communicationnel, et le contenu informationnel associé à cet acte (Godard et Jayez 1996, Milicevic et Polguère 2010). Le caractère hybride des noms comme *déposition*, *témoignage* ou *traduction* s'observe dans des structures co-prédicatives telles que celle présentée en (57), où le premier prédicat pointe la facette OBJET-COGNITIF et le second la facette ÉVÉNEMENT.

- (57) a. [Elle a {lu / écouté}]_{Pred_OBJ-COG} det N (+exp) [effectué(e) par le jeune homme]_{Pred_EVENEMENT}.
 b. Elle a lu {la / le} {**déposition** / **traduction** / **témoignage**} effectué(e) par le jeune homme.

3.5. Les noms hybrides de la classe ÉVÉNEMENT*OBJET-FINANCIER

Les noms hybrides de cette classe complexe dénotent conjointement une action impliquant un transfert d'argent et la somme transférée. La co-prédication en (58a) indique que ces deux sens sont compatibles pour les noms testés en (58b).

- (58) a. [Elle a {effectué}]_{Pred_EVENEMENT} det N (+exp) [modique]_{Pred_OBJ-FIN}
 b. Elle a effectué un(e) {**dépense / donation / emprunt / versement**} modique.

3.6. Les noms hybrides de la classe ÉVÉNEMENT*PHÉNOMÈNE

On classe parmi les noms de PHÉNOMÈNE les noms dénotant des odeurs (ex. *parfum*), des sons (ex. *note*) ou encore des éléments visuels (ex. *lumière*). À côté de cette classe de noms de phénomènes purs, il existe une classe de noms hybrides présentant une facette PHÉNOMÈNE et une facette ÉVÉNEMENT (Cance et Dubois 2005: 22; David 2000: 65). La compatibilité des deux facettes est illustrée dans les co-prédications en (59).

- (59) a. Det N [qui s'est produit à tel moment]_{Pred_EVENEMENT} [a {résonné dans / illuminé} la salle]_{Pred_PHENOMENE}.
 b. La **sonnerie** qui s'est produite inopinément pendant son discours a résonné dans toute la salle.
 c. L'**éclair** qui s'est produit peu après minuit a illuminé toute la salle.

3.7. Les noms hybrides de la classe ÉVÉNEMENT*OBJET-NATUREL

On peut mentionner enfin la classe des noms hybrides dénotant conjointement un événement naturel et les conséquences physiques de cet événement, qui prennent la forme d'un objet naturel. Les co-prédications en (60) illustrent la facette dynamique et la facette d'entité concrète naturelle de noms comme *lésion* et *éboulement*.

- (60) a. Dét N [qui s'est produit à tel endroit]_{Pred_EVENEMENT} [s'étend sur + dimension]_{Pred_OBJ-NAT}.
 b. La **lésion**, qui s'est produite dans une zone délicate, s'étend sur près de trois centimètres carrés.
 c. L'**éboulement**, qui s'est produit dans la nuit, s'étend sur quelques dizaines de mètres.

4. Méthode de classification

Nous présentons ici la méthode adoptée pour classer les noms du français dans une ou plusieurs des classes sémantiques simples ou complexes qui viennent d'être décrites. Cette méthode peut, rappelons-le, s'appliquer aussi bien en langue qu'en discours. Elle peut ainsi être utilisée pour classer les différents sens lexicalisés d'un nom ou pour classer des occurrences de ce nom en corpus. Dans les deux cas, il s'agit bien de classer des formes lexicales spécifiées sémantiquement, ce qui implique que l'identification du sens en jeu ait été opérée en amont.

4.1. Arbre de décision

La méthode de classification consiste à appliquer successivement à un nom, pris dans l'un de ses sens s'il en a plusieurs, les tests permettant d'identifier la classe sémantique à laquelle il appartient¹⁸. L'ordre dans lequel ces tests doivent être appliqués est représenté dans l'arbre de décision dans la figure 1 ci-après.

Les feuilles de l'arbre de décision correspondent aux différentes classes sémantiques retenues. Les nœuds, quant à eux, correspondent soit à l'application d'un test (nœuds bleus) soit à l'identification d'une classe pouvant entrer dans la composition d'une classe complexe (nœuds jaunes) et devant faire l'objet d'une étape d'analyse supplémentaire (cf. *infra*).

L'arbre est construit de manière à respecter l'ordre d'importance des classes (on teste en premier l'appartenance aux classes les plus fréquentes) et de manière à ce qu'à chaque nœud, la réponse "oui" au test conduise le plus directement possible à la classification du nom testé. C'est pourquoi le premier nœud de l'arbre ne correspond pas au test T1, commun aux trois classes ANIMÉ, ARTEFACT et OBJET-NATUREL, mais au test T2, qui distingue uniquement la classe ANIMÉ des autres classes.

La classification proposée a pour ambition de couvrir entièrement le lexique nominal français, cependant, il est inévitable que certains N, qui ne répondent positivement à aucun test, demeurent sans étiquette (cf. feuille "autre" en bas de l'arbre).

4.2. Classification des N relevant de classes complexes

En cas de réponse positive à un test, la classification du nom n'est directe que pour les classes qui ne sont pas suivies d'un astérisque. Celles qui le sont, ARTEFACT*, OBJET-NATUREL* et ÉVÉNEMENT*, sont celles susceptibles d'entrer dans la composition d'une classe complexe (cf. *supra*, section 3). La figure 2 détaille les trois sous-arbres concernés.

La démarche, qui consiste à évaluer si le nom testé appartient à la classe simple ou à l'une des classes complexes associées, est la même dans les trois sous-arbres. On évalue tout d'abord si le nom valide le test révélant l'appartenance à l'autre classe constitutive de la classe complexe. Par exemple, dans le sous-arbre OBJET-NATUREL*, auquel on accède quand un nom a passé le test révélateur de la classe OBJET-NATUREL, on évalue au moyen de T4 si le nom en cours d'analyse est aussi de type ÉVÉNEMENT, puisqu'il existe une classe complexe ÉVÉNEMENT*OBJET-NATUREL (ex. *éboulement*). Le cas échéant, on construit une co-prédication pour déterminer si les deux sens sont compatibles ou non. Si le nom entre dans une co-prédication, il est rangé dans la classe complexe. Si au contraire, on ne trouve pas de co-prédication naturelle, on doit considérer que la forme testée est en fait polysémique, et prendre position sur le sens lexical en cours d'analyse. Enfin, si le nom ne passe aucun test révélateur de l'autre classe formant la classe complexe, il est rangé dans la classe simple.

¹⁸Pour rappel, l'intuition du linguiste sur l'acceptabilité du test appliqué doit, lorsque cela est possible, s'appuyer sur une observation en corpus. On cherchera par exemple en corpus s'il existe des occurrences de « *le loto a eu lieu* » si un doute est possible sur l'acceptabilité de l'application de T4a à *loto*.

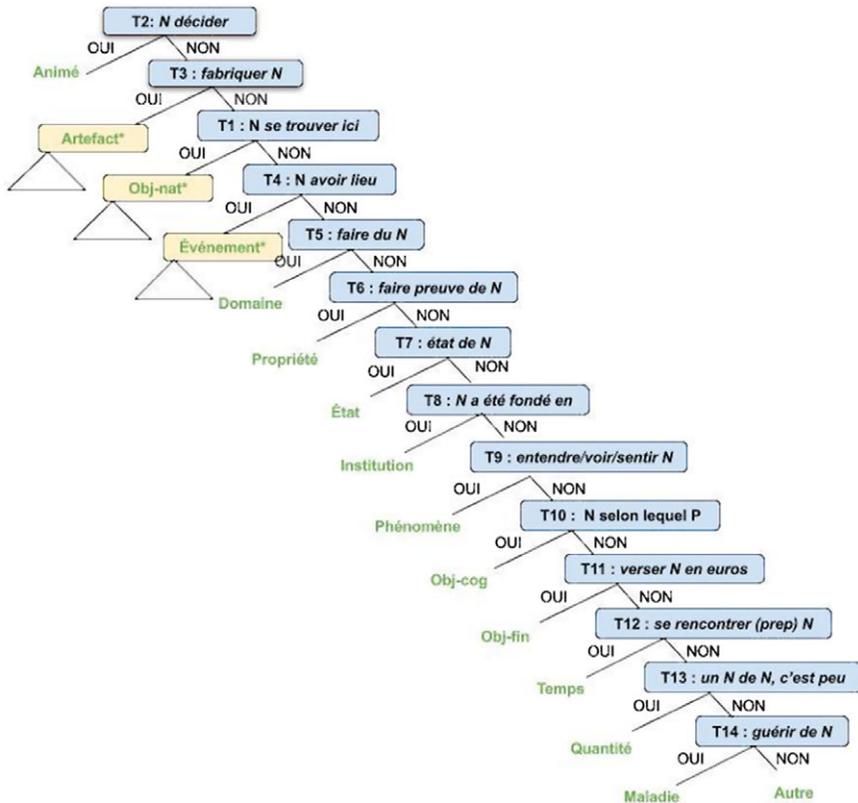


Figure 1: Arbre de décision pour la classification.

4.3. Exemples d'application de la méthode de classification

Les principaux éléments de la méthode de classification ayant été exposés, prenons deux exemples d'application : les noms *étoile* et *discussion* considérés dans leurs emplois en (61).

- (61) a. Des centaines de planètes gravitent autour d'autres **étoiles** que le soleil.
- b. Les **discussions** ont essentiellement porté sur le diamant, principale richesse de la province.

Le nom *étoile* est très polysémique (il peut dénoter, entre autres, un astre, une forme, une distinction, une personne), mais c'est dans son sens premier de corps céleste qu'il est employé en (61a) et c'est donc dans ce sens uniquement qu'il sera considéré ici. La réponse "non" au premier test appliqué (*? L'étoile a décidé de se diriger vers le soleil*) nous conduit à l'application du test suivant, qui donne lui aussi une réponse négative (*? Il a fabriqué une étoile proche du soleil*). Le test suivant, en revanche, s'applique bien au sens d'*étoile* considéré (*L'étoile se trouve à plusieurs années-lumière du soleil et une étoile d'un million de kilomètres de diamètre*), plaçant le nom dans la classe des noms d'OBJET-NATUREL.

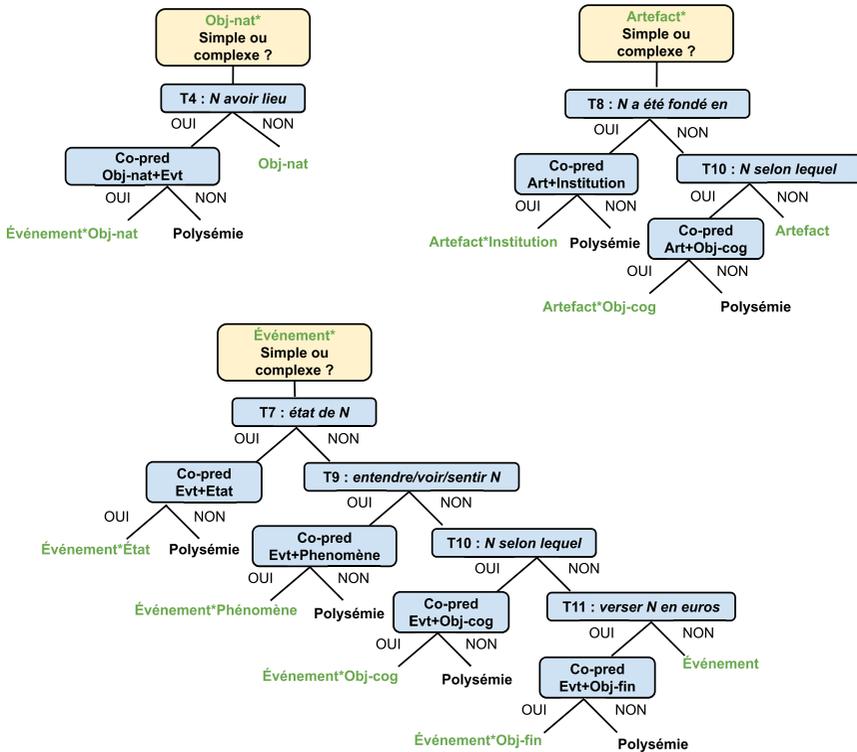


Figure 2: Sous-arbres pour l'identification des classes complexes.

Considérons maintenant le nom *discussion* en (61b). L'application des trois premiers tests (T2, puis T3, puis T1) conduit chaque fois à une réponse négative, indiquant que le nom ne dénote pas une entité concrète. Le nom répond en revanche positivement au test de dynamicité (*la discussion a eu lieu dans son bureau*), conduisant au nœud ÉVÉNEMENT* dans l'arbre. Il faut alors déterminer si *discussion* appartient à la classe simple des ÉVÉNEMENT ou s'il appartient à une classe complexe impliquant le type ÉVÉNEMENT (cf. *supra* figure 2). On applique successivement le test disjonctif de stativité (T7), que le nom testé ne valide pas (? *Pierre ressent de la discussion* ou ? *Pierre est dans un état de discussion*), puis T9, qui n'est pas valide non plus (? *On a entendu une discussion tonitruante* / *Une discussion a retenti dans la cour*), puis T10, qui indique que *discussion* dénote également un OBJET-COGNITIF (*Il écoute les discussions décousues des convives*). Le nom pouvant enfin entrer dans une co-prédication (*il a écouté les discussions intéressantes qui ont eu lieu à l'issue de la rencontre*), on considérera qu'il relève de la classe complexe ÉVÉNEMENT*OBJET-COGNITIF.

Conclusion

Nous avons présenté dans cet article une typologie sémantique à gros grain, visant à couvrir l'ensemble des noms du français, et une méthode de classification de ces noms fondée sur l'application systématique de tests linguistiques. La typologie proposée est constituée de 14 classes simples et de 7 classes complexes, exclusives les unes des autres. La prise en compte des classes complexes, qui n'avait à notre connaissance pas fait jusqu'ici l'objet d'un recensement pour les noms du français, permet de distinguer clairement les différents cas de sens multiples dans le domaine nominal. Les noms homonymiques et polysémiques relèvent d'autant de classes qu'ils ont de sens en langue tandis que les noms hybrides, qui ont des sens distincts, mais compatibles, relèvent de classes complexes.

La méthode de classification proposée repose de manière cruciale sur l'utilisation de tests linguistiques. Inspirés pour la plupart de travaux existants, ils ont été repensés et amendés pour convenir à notre approche de classification générale des noms du français. Pour être révélateur d'une classe sémantique de noms et seulement d'elle, chaque test a été défini avec soin, illustré systématiquement et ses éventuelles contraintes d'utilisation (types de compléments à utiliser, types de déterminants, *etc.*) ont été explicitées chaque fois que nécessaire. Nous avons conscience du caractère potentiellement rigide de ces tests, qui peuvent laisser en dehors de l'analyse certains noms, notamment ceux apparaissant de manière plus naturelle avec des prédicats qui ne correspondent pas à ceux retenus pour les tests. Il est néanmoins possible, selon la visée de l'étude menée, de remodeler un test pour l'élargir ou au contraire le restreindre davantage, tout en gardant à l'esprit la nécessité d'évaluer dans de vastes corpus les implications de ces éventuels changements.

Nous espérons que le résultat de ce travail, issu de notre expérience acquise au fil de différentes campagnes d'annotation des noms du français, pourra s'avérer utile aux linguistes devant recourir à la classification sémantique pour toute étude ayant trait au lexique (sémantique, morphologie, micro-syntaxe)¹⁹. L'objectif essentiel était pour nous de proposer un mode de classification proprement linguistique, accessible aux non-spécialistes, et empiriquement exploitable à grande échelle.

Competing interests. The authors declare none.

Déclaration de conflit d'intérêts. Les auteur-es déclarent que cette étude a été menée indépendamment de toute relation commerciale ou financière qui pourrait constituer un potentiel conflit d'intérêts.

Conflict of interest statement. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

¹⁹La validité de la classification proposée a été testée à l'occasion de l'annotation sémantique d'un échantillon de 501 noms néologiques (Huyghe et al. à paraître). Cette annotation a été réalisée en double aveugle. Le taux d'accord inter-annotateur observé est de 83%. Une évaluation de cet accord prenant en considération l'accord par chance (kappa de Cohen) donne un résultat de 0,77. Ces mesures indiquent un accord substantiel entre annotateurs.

Références

- Asnès, M. et Kupferman, L. (2008). Évènements, prédicats, arguments : quelques points de repère. *Langages*, 169: 7–33.
- Ansambré, J.-C. (1995). Morphologie et représentation événementielle : le cas des noms de sentiment et d'attitude. *Langue française*, 105: 40–54.
- Apothéloz, D. (2008). Entrer quelques instants vs. arriver quelques instants : le problème de la spécification de la durée de l'état résultant. *Verbum*, 30: 199–219.
- Arapinis, A. et Vieu, L. (2015). A plea for complex categories in ontologies. *Applied Ontology*, 10: 285–296.
- Asher, N. (2011). *Lexical Meaning in Context*. Cambridge: Cambridge University Press.
- Asher, N. et Pustejovsky, J. (2006). A type composition logic for generative lexicon. *Journal of Cognitive Science*, 6.1: 1–38.
- Balibar-Mrabeti, A. (1990). Analyse d'adverbes en "dans". *Langue Française*, 86: 65–74.
- Balibar-Mrabeti, A. (dir.) (1995). *Grammaire des sentiments*. = *Langue Française* 105. Paris: Larousse.
- Balvet, A., Barque, L., Condet, M.-H., Haas, P., Huyghe, R., Marin, R. et Merlo, A. (2011). La ressource Nomage. Confronter les attentes théoriques aux observations du comportement linguistique des nominalisations en corpus. *Traitement Automatique des Langues*, 52.3: 1–24.
- Barque, L., Haas, P., Huyghe, R., Tribout, D., Candito, M., Crabbé, B. et Segonne, V. (2020). FrSemCor: Annotating a French Corpus with Supersenses. In: N. Calzolari et al. (dirs.), *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, pp. 5912–5918.
- Beausero, D. (2009). *Syntaxe et sémantique des noms abstraits statifs : des propriétés verbales ou adjectivales aux propriétés nominales*. Thèse de doctorat. Nancy: Université de Nancy.
- Benninger, C. (1999). *De la quantité aux substantifs quantificateurs*. Metz: Université de Metz.
- Berthonneau, A.-M. (1989). *Composantes linguistiques de la référence temporelle. Les compléments de temps, du lexique à l'énoncé*. Thèse d'État. Université Paris 7 – Denis Diderot.
- Cance, C. et Dubois, D. (2005). Dire notre expérience du sonore : nomination et référenciation. *Langue Française*, 188: 15–32.
- Carlson, G. (1977). *Reference to kinds in English*. PhD. University of Massachusetts at Amherst.
- Copestake, A. et Briscoe, T. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics*, 12.1: 15–67.
- Cruse, D.A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Cruse, D. A. (1995). Polysemy and related phenomena from a cognitive linguistic viewpoint. In: P. St. Dizier et E. Viegas (dirs.), *Computational Lexical Semantics*, Cambridge: Cambridge University Press, pp. 33–39.
- Cruse, D. A. (2000). Lexical "facets": between monosemy and polysemy. In: S. Beckmann, P. P. König et T. Wolf (dirs.), *Sprachspiel und Bedeutung: Festschrift für Franz Hundsnurscher zum 60 Geburtstag*, Tübingen: Max Niemeyer Verlag, pp. 25–36.
- Cruse, D.A. (2004). Lexical facets and metonymy. *Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies*, 47: 73–96.
- Daladier, A. (1999). Auxiliation des noms d'action. *Langages*, 135: 87–107.
- David, S. (2000). Certitudes et incertitudes dans les domaines olfactif, gustatif et auditif. *Cahiers du LCPE*, 4: 77–108.
- Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67: 547–619.
- Ehrmann, M. (2008). *Les Entités Nommées, de la Linguistique au TAL. Statut Théorique et Méthodes de Désambiguïsation*. Thèse de doctorat. Paris: Université Paris 7 – Denis Diderot.
- Falkum, I. et Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua*, 157: 1–16.
- Flaux, N. et Stosic, D. (2012). Les noms d'idéalités sont-ils polysémiques ? In : L. de Saussure et A. Rihs (dirs.), *Études de sémantique et pragmatique françaises*. Berne: Peter Lang, pp. 167–190.
- Flaux, N. et Stosic, D. (2014). Les noms d'idéalités et la nominalisation. In : J. Goes, C. Lachet et A. Masset (dirs.), *Nominalisation*. Artois: Artois Presses Université, pp. 19–38.
- Flaux, N. et Van de Velde, D. (2000). *Les noms en français : esquisse de classement*. Paris: Ophrys.
- Fradin, B. (2011). Remarks on state-denoting nominalizations, *Recherches Linguistiques de Vincennes*, 40: 73–99.
- Frisson, S. (2015). About bound and scary books: the processing of book polysemies. *Lingua*, 157: 17–35.

- Gaatone, D.** (2004). Ces insupportables verbes supports : le cas des verbes événementiels. *Linguisticae Investigationes*, 27.2: 239–251.
- Garey, H.B.** (1957). Verbal aspect in French. *Language*, 33: 91–110.
- Giry-Schneider, J.** (1978). *Les nominalisations en français. L'opérateur "faire" dans le lexique*. Genève: Droz.
- Giry-Schneider, J.** (1987). *Les prédicats nominaux en français. Les phrases simples à verbe support*. Genève: Droz.
- Godard, D. et Jayez, J.** (1996). Types nominaux et anaphores : le cas des objets et des événements. *Cahiers Chronos*, 1: 41–58.
- Goossens, V.** (2009). La polysémie des noms d'affect. *Tranel*, 50: 145–158.
- Gotham, M.** (2017). Composing Criteria of Individuation in Copredication. *Journal of Semantics*, 34.2: 333–371.
- Grouin, C., Rosset, S., Zweigenbaum, P., Fort, K., Galibert, O. et Quintard, L.** (2011). Proposal for an Extension of Traditional Named Entities: from Guidelines to Evaluation, an Overview. In: N. Ide, A. Meyers, S. Pradhan et K. Tomanek (dirs.), *Proceedings of the 5th Linguistic Annotation Workshop*. Association for Computational Linguistics, pp. 92–100.
- Gross, G.** (1994). Classes d'objets et description des verbes. *Langages*, 115: 15–30.
- Gross, G.** (1996). Prédicats nominaux et compatibilité aspectuelle. *Langages*, 121: 54–72.
- Gross, G.** (2008). Les classes d'objets. *Lalies*, 28: 111–165.
- Gross, G. et Kiefer, F.** (1995). La structure événementielle des substantifs. *Folia Linguistica*, XXXIX.1–2: 43–65.
- Gross, G. et Vivès, R.** (1986). Les constructions nominales et l'élaboration d'un lexique-grammaire. *Langue Française*, 69: 5–27.
- Gross, M.** (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63: 7–52.
- Grossmann, F. et Tutin, A.** (dir.) (2005). *Sémantique des noms et adjectifs d'émotion*. = LIDIL 32. Grenoble: UGA Éditions.
- Haas, P.** (2009). *Comment l'aspect vient aux noms. Les propriétés aspectuelles des noms à l'épreuve des restrictions de sélection imposées par certaines prépositions*. Thèse de doctorat. Université Lille 3.
- Haas, P., Huyghe, R. et Marín, R.** (2008). Du verbe au nom : calques et décalages aspectuels. In: J. Durand, B. Habert et B. Laks (dirs.), *Actes du 1^{er} Congrès Mondial de Linguistique Française*. SHS Web of Conferences, pp. 2051–2065.
- Haas, P. et Huyghe, R.** (2010). Les propriétés aspectuelles des noms d'activités. *Cahiers Chronos*, 21: 103–118.
- Haas, P. et Jugnet, A.** (2013). De l'existence des prédicats d'achèvements. *Linguisticae Investigationes*, 36.1: 56–89.
- Harris, Z.S.** (1954). Distributional structure. *Word*, 10.2–3: 146–162.
- Hathout, N. et Sajous, F.** (2016). Wiktionnaire's wikicode GLAWified: A workable French machine-readable dictionary. In: N. Calzolari et al. (dirs.), *Proceedings of the 10th Language Resources and Evaluation Conference*. European Language Resources Association, pp. 1369–1376.
- Heyd, S. et Knittel, M.-L.** (2009). Les noms d'activité parmi les noms abstraits : Propriétés aspectuelles, distributionnelles et interprétatives. *Linguisticae Investigationes*, 32: 124–148.
- Huyghe, R.** (2009). *Les noms généraux d'espace en français. Enquête linguistique sur la notion de lieu*. Bruxelles: Duculot.
- Huyghe, R.** (2011). (A)telicity and the mass-count distinction: the case of French activity nominalisations. *Recherches Linguistiques de Vincennes*, 40: 101–126.
- Huyghe, R.** (2014). La sémantique des noms d'action : quelques repères. *Cahiers de Lexicologie*, 105: 181–201.
- Huyghe, R.** (2015). Les typologies nominales : présentation. *Langue française*, 185: 5–27.
- Huyghe, R., Lombard, A., Salvadori, J. et Schwab, S.** (à paraître). Semantic rivalry between French deverbal neologisms in *-age*, *-ion* and *-ment*. In: S. Kotowski et I. Plag (dirs.), *The Semantics of Derivational Morphology: Theory, Methods, Evidence*. Berlin: De Gruyter.
- Jakubicek, M., Kilgarriff, A., Kovar, V., Rychly, P. and Suchomel, V.** (2013). The TenTen corpus family. In: *Proceedings of the 7th International Corpus Linguistics Conference*. Lancaster University, pp. 125–127.
- Kiefer, F.** (1998). Les substantifs déverbaux événementiels, *Langages*, 131: 56–63.
- Kleiber, G.** (1981). *Problèmes de référence : descriptions définies et noms propres*. Paris: Klincksieck.

- Kleiber, G.** (1996). Cognition, sémantique et facettes: une "histoire" de livres et de romans. In: G. Kleiber et M. Riegel (dirs.), *Les Formes du Sens*. Louvain La Neuve: Duculot, pp. 319–331.
- Kleiber, G.** (1999). *Problèmes de sémantique: la polysémie en questions*. Villeneuve d'Ascq: Presses Universitaires du Septentrion.
- Kleiber, G.** (2012). De la dénomination à la désignation : le paradoxe ontologico-désignatif des odeurs. *Langue française*, 174: 45–58.
- Kleiber, G.** (2013). À la recherche de l'intensité. *Langue française*, 177: 63–76.
- Kleiber, G., Benninger, C., Biermann-Fischer, M., Gerhard-Krait, F., Lammert, M., Theissen, A. et Vassiliadou, H.** (2012). Typologie des noms : le critère "se trouver + SP locatif". *Scolia*, 26: 105–129.
- Kleiber, G. et Guillaume, M.** (2011). Sémantique des odeurs. *Langages*, 181: 17–36.
- Labelle, J.** (1986). Grammaire des noms de maladie. *Langages*, 69: 108–125.
- Lammert, M.** (2010). *Sémantique et cognition : les noms collectifs*. Genève: Droz.
- Lammert, M. et Lecolle, M.** (2014). Les noms collectifs en français, une vue d'ensemble. *Cahiers de Lexicologie*, 105.2: 203–222.
- Milicevic, J. et Polguère, A.** (2010). Ambivalence sémantique des noms de communication langagière du français. In : F. Neveu, V. Muni Toke, T. Klingler, J. Durand, L. Mondada et S. Prévost (dirs.), *Actes du 2ème Congrès Mondial de Linguistique Française*. SHS Web of Conferences, pp. 1029–1050.
- Murphy, E.** (2021). *Linguistic Representation and Processing of Copredication*. Phd. University College London.
- Ortega-Andrés, M. et Vicente, A.** (2019). Polysemy and co-predication. *Glossa: a journal of general linguistics*, 4.1: 1–23.
- Petrolito, T. et Bond, F.** (2014). A survey of wordnet annotated corpora. In: H. Orav, C. Fellbaum et P. Vossen (dirs.), *Proceedings of the Seventh Global WordNet Conference*. Tartu: University of Tartu Press, pp. 236–243.
- Piñón, C.** (1997). Achievements in an Event Semantics. In: A. Lawson et E. Cho (dirs.), *Proceedings of Semantic and Linguistic Theory 7*. Ithaca (N.Y.): Cornell University Press, pp. 273–296.
- Piñón, C.** (1999). Durative adverbials for result states. In: S. Bird, A. Carnie, J. D. Haugen, et P. Norquest (dirs.), *Proceedings of the 18th West Coast Conference on Formal Linguistics*. Somerville (Mass.): Casadilla Press, pp. 420–433.
- Pustejovsky, J.** (1995). *The Generative Lexicon*. Cambridge (Mass.): MIT Press.
- Rastier, F.** (2002). Valeur saussurienne et valeur monétaire. *L'information grammaticale*, 95: 46–49.
- Schäfer, R.** (2015). Processing and querying large web corpora with the COW14 architecture. In: P. Banski, H. Biber, E. Breiteneder, M. Kupietz, H. Längen et A. Witt (dirs.), *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora*. Mannheim: Institut für Deutsche Sprache, pp. 28–34.
- Schäfer, R. and Bildhauer, F.** (2012). Building large corpora from the web using a new efficient tool chain. In: N. Calzolari et al. (dirs.), *Proceedings of the 8th Language Resources and Evaluation Conference*. European Language Resources Association, pp. 486–493.
- Tribout, D., Barque, L., Haas, P. et Huyghe, R.** (2014). De la simplicité en morphologie. In: F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefer et S. Prévost (dirs.), *Actes du 4ème Congrès Mondial de Linguistique Française*. SHS Web of Conferences, pp. 1879–1890.
- Tutin, A., Novakova, I., Grossman, F. et Cavalla, C.** (2006). Esquisse de typologie des noms d'affect à partir de leurs propriétés combinatoires. *Langue française*, 150: 32–49.
- Van de Velde, D.** (1995). *Le spectre nominal : des noms de matière aux noms d'abstractions*. Louvain-Paris: Peeters.
- Van de Velde, D.** (1997). Un dispositif linguistique propre à faire entrer certaines activités dans des taxinomies : "Faire + du + Nom d'activité". *Revue de Linguistique Romane*, 243–244: 369–395.
- Vendler, Z.** (1967). *Linguistics in Philosophy*. Ithaca (N.Y.): Cornell University Press.
- Vigier, D.** (2003). Les syntagmes prépositionnels en "en N" détachés en tête de phrase référant à des domaines d'activité. *Linguisticae investigationes*, 26.1: 97–122.
- Zipf, G. K.** (1949). *Human behavior and the principle of least effort*. Cambridge (Mass.): Addison-Wesley Press.

ANNEXES : liste des tests

T1 → T1a ET T1b (sous la forme b1, b2 ou b3)

a. *Dét N se trouve* + complément de localisation spatiale

ET

b1. *Dét N de* + mesure [pour les N comptables]

b2. mesure + *de N* [pour les N massifs]

b3. *Dét N de Dét.card N₂* [pour les N collectifs]

T2 → T2a ET T2b ET T2c

a. *Dét N {décide / choisit} de SV*

ET

b. *Dét N + SV {volontairement / délibérément / consciemment}*

ET

c. *Dét N a beaucoup mangé*

T3 → T3a OU T3b OU T3c

a. *N_[humain] a {fabriqué / déchiré} Dét N {à tel endroit / à tel moment}*

OU

b. *N_[humain] a construit Dét N + Compl de matière + Adj de couleur*

OU

c. *N_[humain] a confectionné Dét N + Compl d'arôme*

T4 → T4a OU T4b

a. *Dét N {a eu lieu / s'est produit} {à tel moment / à tel endroit}*

OU

b. *N_[humain] {a procédé à / a accompli / a effectué} Dét N (+ expansion facultative)*

T5 → T5a OU T5b

a. *N_[humain] fait {du / de la} N*

OU

b. *N_[humain] a reçu une formation dans le domaine {du / de la} N*

T6 → T6a OU T6b OU T6c

a. *N₀ est d'un(e) grand(e) N*

OU

b. *N_[humain] a fait preuve {de N / d'un(e) N + expansion}*

OU

c. *Cela m'a altéré Dét N*

T7 → T7a OU T7b

a. *N_[humain] {ressent / éprouve} Dét N (+expansion)*

OU

b. *N_[unanimité] est dans un état de N (+expansion)*

T8 → T8a OU T8b

a. *Dét N a été fondé en telle année*

OU

b. *N_[humain] a été nommé à la tête de (Dét) N*

T9 → (T9-a1&a2) OU (T9-b1&b2) OU (T9-c1&c2)

a1. *On a entendu un N (tonitruant / strident / mélodieux / aigu / grave)*

& a2. *Un N (a retenti / a résonné / s'est propagé) (au fond du couloir / dans la cour)*

OU

b1. *On a vu un N (aveuglant / éblouissant / blafard / pâle)*

& b2. *Det N (a resplendi / s'est propagé / a scintillé / a jailli) au fond du couloir*

OU

c1. *On a senti un N (nauséabond / âcre / enivrant / tenace / entêtant)*

& c2. *Det N (s'est répandu / persiste / embaume) (dans) la pièce*

T10 → T10a OU T10b OU T10c

a. *Dét N selon lequel P est difficile à admettre*

OU

b. $N_{[humain]}$ *écoute Dét_[plur] N décousus de Vincent*

OU

c. $N_{[humain]}$ *a écrit Dét N {intéressant / pertinent}*

T11 → T11a OU T11b OU (T11-b1 & T11-b2)

a. *Il a versé Dét N en euros*

OU

b. *Il a obtenu Dét_[indef, sing] N modique*

OU

c1. *Dét N est la monnaie de tel pays*

& c2. *Quel est le taux de change de Dét N?*

T12 → T12a OU T12b

a. *Ils se sont rencontrés {N / courant N / Dét N / Dét.card N durant / à N}*

OU

b. *Attendez Dét_[indef, sing] bref N*

T13 → (T13-a1 & T13-a2) OU T13b

a1. *Dét {courant / poids / lumière / intensité / tension / distance / période / vitesse} de Dét.card N, c'est {peu / beaucoup}*

& a2. *Dét N_0 {mesure / fait / pèse / dure} Dét.card N*

OU

b. *Dét N de N_2 , c'est (peu / beaucoup).*

où N_2 appartient à la liste : *farine, huile, sable, messages, images, cheminées, projectiles, juristes*

T14 → T14a ET T14b ET T14c

a. *Dét N {est / n'est pas} une maladie contagieuse*

ET

b. $N_{[humain]}$ *a guéri de Dét N*

ET

c. *Le médecin m'a diagnostiqué Dét_[indef, sing] N*