

Huyghe Richard & Wauquier Marine. 2021. Une étude distributionnelle des noms d'agent en *-ant*, *-eur*, *-ien*, *-ier* et *-iste*. *Verbum* 43, 13-40.

## UNE ÉTUDE DISTRIBUTIONNELLE DES NOMS D'AGENT EN *-ANT*, *-EUR*, *-IEN*, *-IER* ET *-ISTE*

**Richard HUYGHE**  
Université de Fribourg

**Marine WAUQUIER**  
CLLE, CNRS & Université Toulouse Jean Jaurès

### RÉSUMÉ

*Dans cet article, nous comparons différents suffixes formateurs de noms d'agent en français. Nous nous interrogeons sur l'existence de corrélations entre la construction morphologique de ces noms et leurs spécificités sémantiques. Nous adoptons une approche distributionnelle qui nous permet d'étudier à grande échelle les noms d'agent et de mettre en évidence certaines tendances sémantiques associées aux différents suffixes.*

### ABSTRACT

*In this paper, we compare several suffixes used to form agent nouns in French. We investigate whether the morphological properties of agent nouns correlate with semantic specificities. We adopt a distributional semantics approach that allows for a large-scale study of agent nouns, and for the identification of semantic tendencies associated with the different agentive suffixes.*

### INTRODUCTION<sup>1</sup>

Les noms d'agent en français, i.e. les noms décrivant des entités qui effectuent une action, sont morphologiquement divers. On trouve parmi eux des noms dérivés (*vendeur*), mais aussi des composés (*aide-soignant*), des noms en relation de conversion avec un verbe (*judge*) ou un adjectif (*criminel*), des noms tronqués (*indic*) et/ou rédupliqués (*nounou*), des amalgames (*entrepreneute*, à partir d'*entrepreneur* et *internaute*), des acronymes (*dircab*),

---

<sup>1</sup>Nous remercions les deux relecteurs anonymes pour leurs remarques sur une précédente version de cet article.

des complexes non construits (*ivrogne*)<sup>2</sup>, voire des noms morphologiquement simples (*médecin*). Lorsqu'ils sont dérivés, les noms d'agent (Nag) résultent de constructions variées, en ce qui concerne la catégorie grammaticale et du type sémantique de la base sélectionnée, ainsi que la forme du nom construit. Il existe différents suffixes agentifs, dont la productivité est très variable — si les Nag en *-eur* sont très abondants, d'autres, comme les Nag en *-ard* (*motard*, *thésard*) et en *-on* (*forgeron*, *maton*), sont beaucoup plus rares, voire résiduels. Certains suffixes agentifs du français ont fait l'objet d'études spécifiques. C'est le cas notamment de *-eur*, *-ier*, *-ien*, *-ant*<sup>3</sup> et *-iste* (cf. Dubois 1962, Winther 1975, Corbin & Corbin 1991, Lignon 2000, Anscombe 2001, Fradin & Kerleroux 2003, Roché 2004, 2011, Rosenberg 2008, Roy & Soare 2012, Cartoni & Namer 2012, Huyghe & Tribout 2015, i.a.). Ces suffixes sont souvent étudiés isolément, parfois comparés deux à deux (notamment dans le cas de *-eur* et *-ant*, ou de *-ien* et *-iste*), mais rares sont les études qui étendent la comparaison à plus de deux formes.

Dans cet article, nous nous proposons d'étudier conjointement les Nag en *-ant/-ante*, *-eur/-euse/-rice*, *-ien/-ienne*, *-ier/-ière* et *-iste* (notés respectivement *-ant*, *-eur*, *-ien*, *-ier*, *-iste*). Il s'agit de savoir si les disparités de construction morphologique de ces noms sont corrélées à des différences de sémantique fine. Nous nous demandons en particulier si les suffixes agentifs sont associés à différents modes d'élaboration sémantique de l'agentivité et à la description de types d'agents distincts. Pour répondre à cette question, nous nous appuyons sur certains outils de sémantique distributionnelle développés en linguistique informatique, qui nous permettront d'évaluer empiriquement la similarité sémantique entre les différents Nag. Nous postulons en effet que des différences de distribution récurrentes entre des ensembles de Nag différemment suffixés sont le fait de

---

<sup>2</sup>Corbin (1987 : 459) définit les mots complexes non construits comme des mots dont « la structure interne et le sens ne sont que partiellement superposables », et dont l'analysabilité morphosémantique n'est donc pas complète. Dans le cas d'*ivrogne*, on peut identifier une base formelle et sémantique, *ivre*, mais la chaîne de caractères *-ogne* n'est l'exposant d'aucune règle de dérivation.

<sup>3</sup>Le statut morphologique de *-ant* est incertain et débattu dans les travaux existants. Les Nag en *-ant* (*combattant*, *manifestant*) se présentent *a priori* comme des convertis de formes verbales participiales (Anscombe 2003), mais certains auteurs argumentent en faveur d'une analyse suffixale (Dubois 1962, Winther 1975, Ulland 1993). On peut noter que l'hypothèse de la conversion ne convient pas aux Nag en *-ante*, et conduit à la distinction, peu satisfaisante théoriquement, entre deux constructions morphologiques distinctes pour les Nag masculins et féminins en *-ant* et *-ante* (respectivement conversion et dérivation). Arrêter une position dans ce débat demanderait néanmoins de plus amples développements. Par défaut, nous intégrons dans notre étude les Nag en *-ant* et *-ante*, en nous fondant sur le fait que même en cas de conversion, l'orientation de dérivation du verbe au nom est établie, et que donc les noms concernés sont effectivement des Nag construits. Par commodité, nous référerons ici collectivement à l'ensemble des Nag étudiés comme à des noms suffixés (dans une conception ouverte de cette catégorie).

types sémantiques distincts, et que des différences distributionnelles plus ou moins marquées révèlent des écarts sémantiques plus ou moins importants entre les suffixes concernés.

L'article est organisé comme suit. Dans la première partie, nous reviendrons sur la définition et la construction morphologique des Nag en français. Nous présenterons dans la deuxième partie notre méthodologie et les échantillons de Nag utilisés dans l'étude. Les troisième et quatrième parties seront consacrées respectivement aux résultats de l'analyse distributionnelle et à leur discussion.

## **1. LES NOMS D'AGENT EN FRANÇAIS**

Dans cette section, nous présentons brièvement les questions liées à la définition de la notion d'agent et à son application dans le domaine lexical. Nous nous interrogeons ensuite sur la diversité des suffixes agentifs en français, et sur la multiplicité des relations entre forme et sens qu'ils instancient.

### **1.1. L'agentivité en question**

Définir l'agentivité est réputé difficile. Comme l'observe Dowty (1991 : 553), « [Agent] is one of most frequently cited roles, and it is in some sense a very intuitive role, but it is one of the hardest to pin down ». De fait, la notion d'agent reçoit dans la littérature des interprétations variées (cf. Gruber 1967, Fillmore 1968, Cruse 1973, DeLancey 1984, Schlesinger 1989, Kipper Schuler 2005, i.a.). Les conceptions divergent notamment quant à l'animéité et l'intentionnalité de l'agent, voire la dynamicité du prédicat en jeu. Selon les conditions requises pour les deux premiers critères, les sujets dans les exemples (1) feront l'objet d'analyses différentes.

- (1) a. Pierre a assassiné sa femme.
- b. Le sida a tué des millions de personnes.
- c. Pierre a tué sa femme.

*Pierre* dans (1a) dénote un agent, car il vérifie toutes les conditions nécessaires que l'on pourrait poser dans la définition de l'agentivité : le procès décrit est dynamique, le sujet est animé et dans tous les cas intentionnel. Le sujet de (1b) par contre ne sera pas analysé comme un agent (mais plutôt comme une simple cause) si l'intentionnalité ou l'animéité sont requises. Quant au sujet de (1c), il vérifie la condition d'animéité, mais l'intentionnalité dépend de circonstances extralinguistiques, et ne sera dès lors pas systématiquement analysé comme un agent si l'intentionnalité est nécessaire. L'analyse lexicale de *tuer* comme un verbe qui attribue le rôle d'agent dépend donc fondamentalement de la définition choisie. Or selon Gruber, un verbe agentif est un verbe « whose subject refers to an animate object which is

thought of as the wilful source or agent of the activity described by the verb » (1967 : 943), tandis que selon Schlesinger, « it is generally recognized that intention is not a necessary feature of agentivity, for after all we often do things accidentally » (1989 : 194). Selon Van Valin & LaPolla (1997), le nombre de verbes « agentifs », entendus comme des verbes dont le sujet est nécessairement intentionnel, peut être très peu élevé dans certaines langues.

Nous considérerons ici que le rôle d'agent implique *a minima* celui d'effectuateur (i.e. d'élément qui déploie l'énergie nécessaire à la réalisation d'un procès)<sup>4</sup> et que dans le cas canonique, l'agent est animé et intentionnel, sans que ces deux propriétés soient nécessaires<sup>5</sup>. Le point crucial est que la définition retenue puisse convenir à la fois aux analyses discursives et lexicales. Notre objectif par ailleurs est de définir les rôles sémantiques indépendamment des traits ontologiques, de manière à ce que l'assignation de rôles et les restrictions sélectionnelles soient clairement distinguées dans l'analyse des prédicats.

Le fait de ne pas impliquer de condition stricte d'animéité pourrait poser problème s'agissant de la distinction entre agents et instruments, ces deux catégories étant souvent définies par leur animéité et non-animéité respectives. La classification de certaines entités non animées à la limite entre instrumentalité et agentivité (ex. les engins automatisés) pourrait s'avérer délicate. Nous tâcherons de maintenir ici une distinction claire en définissant les instruments comme des entités qui sont fondamentalement utilisées par d'autres entités pour accomplir des actions — ce qui inclut les tracteurs, répondeurs et autres minuteurs. Par ailleurs, impliquer une condition d'animéité pour les agents créerait inévitablement un angle mort pour les forces naturelles et les agents chimiques ou abstraits qui effectuent des actions de manière autonome. Un rôle complémentaire serait nécessaire, et se poserait de nouveau la question de savoir s'il faut distinguer deux structures d'assignation de rôles dans les cas où un prédicat se construit à la fois avec des effectuateurs animés et non animés (ex. *Marie a ouvert la fenêtre* vs *Le vent a ouvert la fenêtre*).

---

<sup>4</sup>Une définition du rôle 'effector' est donnée par Van Valin & LaPolla (1997 : 118) comme "the participant that brings something about, but there is no implication of its being volitional or the original instigator". De la même manière, Cruse (1973 : 21) définit le trait 'agentive' (par opposition à 'volitive', 'effective' et 'initiative') comme "referring to an action performed by an object which is regarded as using its own energy in carrying out the action".

<sup>5</sup>Bien qu'elle soit de structure prototypique, la définition retenue n'est pas comparable à celle que formule Dowty (1991) en termes de "proto-role" agentif. En effet, Dowty inclut parmi les traits définitionnels des proto-agents la perception, ce qui lui permet d'analyser comme proto-agent le sujet de phrases telles que *John knows/believes/is disappointed at the statement* et *John sees/fears Mary*. Les sièges et les expérienceurs sont exclus de la définition proposée ici, car nous présupposons que les agents interviennent dans des situations dynamiques.

## 1.2. Construction morphosémantique des noms d'agent

Les rôles sémantiques sont initialement conçus pour rendre compte des relations établies en contexte, et non pour être exploités dans l'analyse des unités lexicales. Dans l'exemple (2) ci-dessous, *ma soeur* dénote l'agent et *le cambrioleur* le patient.

(2) Ma sœur a assommé le cambrioleur.

L'emploi des rôles sémantiques dans l'analyse lexicale demande que l'on fasse abstraction des relations de discours. Au niveau lexical, des noms peuvent être catégorisés comme relevant d'un rôle donné s'ils décrivent leurs référents comme instanciant ce rôle dans un procès intrinsèquement spécifié. Dans une telle perspective, *soeur* ne sera pas catégorisé comme Nag mais plutôt comme nom relationnel, contrairement à *cambrioleur*, qui est en soi un Nag et non un nom de patient.

La catégorisation comme Nag requiert ainsi que deux conditions soient vérifiées : (i) la présence d'un prédicat d'action dans la structure sémantique du nom, et (ii) la description par le nom de l'agent de cette action. Or ces conditions ne sont pas faciles à établir. Une manière de garantir l'existence d'un prédicat d'action dans la structure sémantique des Nag est de définir les Nag comme des noms déverbaux qui dénotent l'argument agentif de leur base. De fait, le rôle d'agent est régulièrement mobilisé pour caractériser sémantiquement l'output des règles de construction nominale qui opèrent sur des verbes attribuant le rôle d'agent (cf. Booij 1986, Kerleroux 2004, Rosenberg 2008, Roy & Soare 2012, i.a.).

Cette définition présente le double avantage de fournir un principe explicatif de la construction morphosémantique de certains noms déverbaux, et de cantonner le type agentif à un ensemble de noms identifiés formellement. Elle a toutefois l'inconvénient de scinder des classes sémantiquement homogènes. Parmi les noms sous (3a) et (3b), sémantiquement proches, seuls les premiers devraient être catégorisés comme Nag, puisqu'ils se distinguent des seconds par leur construction à partir de verbes d'action.

(3) a. sculpteur, guérisseur, confiseur, rédacteur, manifestant,  
protestataire  
b. artiste, médecin, chocolatier, scribe, gréviste, émeutier

Afin de pallier au risque d'une définition non opératoire sémantiquement, et pour rendre compte des similarités de profil distributionnel des noms tels que (3a) et (3b) (Huyghe & Wauquier à paraître), il faut envisager une définition élargie des Nag, qui ne se limite pas aux noms déverbaux, ni même aux noms dérivés d'une base de sens actionnel.

Cette définition élargie implique cependant qu'il existe diverses constructions morphosémantiques de l'agentivité, et que dans le cas des Nag dérivés, le lien d'agentivité avec la base repose sur différentes relations sémantiques (cf. Huyghe à paraître). Lorsque la base ne décrit pas une action, le prédicat dynamique doit être directement élaboré au niveau nominal. Des bases de types sémantiques différents (artefact, contenu informationnel, domaine, etc.) seront ainsi diversement mobilisées dans l'élaboration du composant actionnel, en induisant des prédicats de natures différentes. Par exemple, les bases qui dénotent des artefacts impliqueront dans le sémantisme des Nag dérivés un prédicat de production (*chocolatier*) ou d'instrumentalisation (*bagagiste*), tandis que les bases qui dénotent des domaines impliqueront un prédicat de pratique intellectuelle (*historien*). On peut se demander si de telles disparités n'ont pas un effet sur le type agentif décrit.

### 1.3. Diversité des suffixes agentifs

La diversité des suffixes agentifs s'observe dans de nombreuses langues (cf. par exemple Andrews 1996 pour le russe, Booij 2002 pour le néerlandais, Luschützky 2011 pour le slavon d'église, Bauer *et al.* 2013 pour l'anglais). L'existence de différents suffixes agentifs dans une langue donnée soulève la question de leur concurrence et de leur possible distinction. Les suffixes peuvent se répartir en fonction de différents facteurs. Ceux-ci peuvent être de nature phonologique : par exemple, Roché (1997) soutient qu'en français une contrainte de non-redoublement du yod bloque la formation de noms comme *\*camionnier* et *\*avionnier*, au profit de noms construits en *-eur* (*camionneur* et *avionneur*). De même, Lignon (2000) montre que, pour des raisons d'euphonie, *-ien* se substitue à *-iste* dans la construction de noms de scientifique lorsque la base se termine par une sifflante (ex. *mathématicien*, *physicien* vs *\*mathématiciste*, *\*physiciste*). Des contraintes de productivité et d'évolution diachronique des suffixes peuvent également jouer un rôle dans leur distribution. Selon Lignon (2007), les Nag en *-ien* qui dénotent des spécialistes et dont la base ne se termine pas par une sifflante (ex. *grammairien*, *historien*) ont été formés dans une période où *-ien* était plus productif que *-iste*. À cet égard, une consultation du site Neoveille<sup>6</sup>, qui répertorie les néologismes dans la presse francophone, suggère que *-eur* (ex. *prestateur*, *instagrammeur*, *canyonneur*), *-ien* (ex. *cogniticien*, *parapharmacien*, *aromaticien*) et *-iste* (ex. *plâtriste*, *neuroscientiste*, *parapentiste*) pourraient être actuellement plus productifs que *-ant* et *-ier*, pour lesquels aucun néologisme n'est recensé.

Il se peut également que l'existence de différents suffixes soit corrélée à des différences de réalisation syntaxique. Roy et Soare (2012), par exemple,

---

<sup>6</sup><https://tal.lipn.univ-paris13.fr/neoveille/html/login.php?action=login>

notent que *-ant* se distingue de *-eur* par la possibilité de dériver d'une base verbale inaccusative (*arrivant* vs *\*arriveur*) ou réflexive (*défendant* vs *défenseur*)<sup>7</sup>. Les différences entre suffixes agentifs peuvent encore être de nature sémantique, même si cet aspect est rarement directement traité dans les travaux existants. Les Nag en *-iste* et en *-ien* sont généralement décrits comme des noms de spécialiste (Lignon 2000, Roché 2011, Cartoni et Namer 2012). Roché (2004) remarque que le suffixe *-ier* permet de former des noms de métier dont la base est le plus souvent un nom d'objet exprimant un proto-patient (*horloger, gondolier*), et plus marginalement un nom d'activité (*couturier, cuisinier*). Seuls Cartoni et ses collègues (2015) examinent conjointement les Nag en *-eur, -ien, -iste, -ier*, dans une étude comparée avec les Nag italiens suffixés en *-aio, -ario, -iano, -iere, -ista* et *-ore*. Les suffixes français employés dans la traduction des Nag italiens sont analysés selon le sous-type agentif de ces derniers, ce qui peut indirectement révéler certaines spécificités des suffixes en question. Il apparaît par exemple que les Nag français qui traduisent les Nag italiens d'activité manuelle et de vente sont principalement des noms en *-eur* et en *-ier*, que les agents d'activités artistiques et culturelles sont dénotés principalement par des Nag en *-eur* et *-iste*, ou que la description d'activités illicites incombe principalement aux Nag en *-eur*. On observe en filigrane qu'aucun suffixe ne se cantonne à un type d'agent spécifique.

Certaines distinctions internes à des classes morphologiques de Nag sont par ailleurs décrites. C'est le cas en particulier pour les Nag en *-eur*. À la suite de Benveniste (1948), on distingue entre les noms qui dénotent des agents statutaires, caractérisés indépendamment de leur participation à un événement spécifique (ex. *inspecteur, déménageur, tatoueur, coiffeur, professeur*), et les noms qui dénotent des agents occasionnels, en vertu de la réalisation d'une occurrence particulière d'une action (ex. *agresseur, expéditeur, fondateur, ravisseur, vainqueur*). Les premiers peuvent s'employer comme prédicats nus dans des phrases attributives (4), tandis que les seconds peuvent s'employer avec une structure argumentale renvoyant aux participants d'une action spécifique (5).

- (4) a. Pierre est {inspecteur / déménageur / tatoueur}.  
 b. ??Pierre est {agresseur / expéditeur / fondateur}.

---

<sup>7</sup>Roy et Soare (2012) jugent par ailleurs que les noms en *-ant*, contrairement à ceux en *-eur*, n'expriment pas le rôle d'agent du verbe de base. Les éléments apportés à l'appui de cette thèse sont toutefois contestables et non vérifiés dans les faits. Par exemple, le jugement d'agrammaticalité pour des exemples comme *des enseignants obstinés, un votant déterminé* semble pour le moins discutable — fût-ce dans l'interprétation visée, i.e. orientée vers l'action, et non prédicative.

- (5) a. l'agresseur de la vieille dame, l'expéditeur de ce colis, le fondateur de cette ville  
 b. ??l'inspecteur de l'appartement, ??le déménageur de ces meubles, ??le tatoueur de ce motif

Une troisième catégorie peut être envisagée pour les noms qui dénotent des agents dispositionnels, décrits en fonction de tendances à agir ou à se comporter d'une certaine manière (ex. *séducteur*, *bluffeur*, *bouquineur*, *râleur*, *bagarreur*, cf. Huyghe & Tribout 2015). Ces Nag se construisent difficilement dans les tournures précédentes (6a-b), mais ils ont la particularité de pouvoir s'employer avec des adjectifs (de taille par exemple) dans une lecture non intersective (Anscombe 2001, Roy & Soare 2012), comme dans (6c).

- (6) a. ??Pierre est {séducteur / bluffeur / bouquineur}<sup>8</sup>.  
 b. ??le séducteur de la jeune fille, ??le bluffeur du client, ??le bouquineur de ce livre  
 c. un grand séducteur, un énorme bluffeur, un gros bouquineur

On peut se demander si cette tripartition établie pour les Nag en *-eur* se retrouve pour les autres classes de Nag suffixés.

Il est par ailleurs couramment signalé que les suffixes utilisés dans la formation des noms d'agent ne se cantonnent pas à la formation agentive. Or les multiples capacités de construction des différents suffixes agentifs ne sont pas identiques. Il est connu que *-eur* peut former des noms d'instrument, tels que *réfrigérateur* et *aspirateur*. C'est aussi le cas de *-ier* (*gaufrier*), mais ce dernier a en outre la possibilité de dériver des noms de récipient (*sucrier*), d'arbre fruitier (*cerisier*), de lieu (*rizière*), d'ensemble (*pierrier*), etc. (cf. Corbin & Corbin 1991, Roché 2003). Les noms en *-ant*, pour leur part, peuvent dénoter des substances en tant que moyens d'action (*désherbant*, cf. Knittel 2017), les noms en *-ien* des habitants (*Malien*, cf. Lignon 2000), et les noms en *-iste* des adeptes de doctrines (*marxiste*, cf. Roché 2011, Lignon & Namer 2018). En raison de ces différences de capacité dérivationnelle, la concurrence entre *-ant*, *-eur*, *-ien*, *-ier*, *-iste* n'est, au mieux, que partielle. Les différences relevées contribuent à distinguer les suffixes et à justifier l'existence de chacun dans le système dérivationnel du français. Elles peuvent également éclairer certaines spécificités des interprétations agentives de chaque suffixe. On peut en effet se demander s'il existe un lien sémantique entre les différentes interprétations associées à chaque suffixe. Par exemple, le fait que *-ier* construise généralement des noms qui dénotent des fournisseurs de l'entité décrite par la base peut être mis en relation avec le fait que les Nag en *-ier* réfèrent à des fabricants ou à des vendeurs d'objets manufacturés. On

<sup>8</sup> Le jugement d'acceptabilité concerne ici l'emploi nominal, et non adjectival, des formes testées.

peut de la même manière concevoir pour *-iste* un lien sémantique entre la construction de noms d'adepte et de spécialiste, fondé sur la présupposition de connaissances liées à un domaine spécifique. Nous n'étudierons pas ici les interprétations non agentives des suffixes examinés, mais celles-ci pourront faire l'objet d'une future recherche, en lien avec l'analyse des relations multiples entre forme et sens établies dans les procédés de dérivation.

Dans la suite de ce travail, nous nous concentrerons sur les interprétations agentives de *-ant*, *-eur*, *-ien*, *-ier* et *-iste*, en adoptant une approche distributionnelle. Nous nous demanderons si une analyse quantitative fondée sur la distribution des différents Nag peut confirmer certaines spécificités sémantiques relevées dans la littérature, et éventuellement en faire apparaître d'autres.

## **2. MÉTHODOLOGIE**

Dans cette section, nous présentons notre dispositif d'analyse distributionnelle. Nous rappelons d'abord les fondements théoriques à partir desquels nous menons cette analyse. Nous détaillons ensuite la méthode de sélection des Nag utilisés dans l'étude, et examinons les propriétés des échantillons de Nag ainsi constitués.

### **2.1. Principes de sémantique distributionnelle**

La sémantique distributionnelle s'appuie sur l'hypothèse distributionnelle (Harris 1954, Firth 1957), qui stipule que le sens des mots peut être étudié au travers de leur distribution. Selon cette approche, la différence de sens entre deux mots est corrélée à la différence de contexte d'apparition, et leur proximité sémantique peut être évaluée d'après la similarité de leur distribution en corpus.

Du fait de leur accessibilité et de leurs performances, les outils de sémantique distributionnelle, tels que Word2Vec (Mikolov *et al.* 2013a), sont largement utilisés en linguistique dans des travaux touchant à la sémantique, la syntaxe ou la morphologie. Des thématiques aussi variées que le changement sémantique, la compositionnalité, la classification verbale, la sélection argumentale, ou la différenciation des affixes ont été étudiées à l'aune de cette approche computationnelle (Lenci 2018, Boleda 2020).

Word2Vec construit à partir d'un corpus donné une représentation géométrique du sens des mots, sous forme de vecteurs qui agrègent l'information distributionnelle de chaque mot du corpus. Le modèle statistique résultant, appelé modèle ou espace vectoriel, peut être utilisé pour quantifier la proximité distributionnelle, et par extension la similarité sémantique, des mots. Les vecteurs des mots partageant des contextes similaires sont proches dans le modèle construit, et réciproquement, les mots dont les vecteurs sont proches dans l'espace vectoriel sont considérés comme sémantiquement proches. Cette proximité est quantifiée au moyen de la distance cosinus entre

les vecteurs, et traduite par un score allant de 0 (proximité nulle) à 1 (identité des vecteurs). Les modèles peuvent être interrogés pour déterminer les voisins distributionnels d'un mot donné, i.e. les mots dont les vecteurs sont les plus proches de celui du mot ciblé. Du fait de leur nature mathématique, les vecteurs peuvent par ailleurs être combinés pour reproduire la compositionnalité sémantique (Mikolov *et al.* 2013b).

Word2Vec permet de comparer sur le plan distributionnel des items deux à deux. Par exemple, les noms *danseur* et *pianiste* ont un score de proximité de 0,721 dans le corpus *Wikipédia 2018* (que nous décrivons plus bas), alors que *danseur* et *pomme* ont un score de 0,173. La comparaison de ces scores suggère donc que *danseur* est sémantiquement plus proche de *pianiste* que de *pomme*. Nous pouvons par ailleurs comparer leurs voisins respectifs dans l'espace vectoriel. Ainsi, les plus proches voisins de *danseur* dans *Wikipédia 2018* sont *chorégraphe*, *acrobate* et *ballerine*, alors que les plus proches voisins de *pianiste* sont *violoniste*, *violoncelliste* et *compositeur*. Une telle démarche ne permet toutefois pas d'avoir une vision d'ensemble du sens associé aux Nag en *-eur* et en *-iste* en tant que classes sémantiques. Pour ce faire, nous exploitons la compositionnalité des vecteurs. En nous inspirant de Kintsch (2001) et Wauquier *et al.* (2020), nous faisons l'hypothèse que les traits sémantiques partagés par les mots d'une même classe peuvent être approchés au travers de la représentation distributionnelle de l'ensemble des mots de cette classe. Plus précisément, nous construisons une représentation distributionnelle unique (appelée *barycentre*) qui agrège les propriétés distributionnelles des mots en question, en faisant la moyenne de leurs vecteurs. Ce barycentre est considéré comme représentatif des mots initialement ciblés (que nous appellerons *amorces*). Dans le cas des Nag, nous construisons un barycentre pour chaque classe de Nag définie par un suffixe donné. Nous évaluons ensuite leurs propriétés en analysant les 100 plus proches voisins de ces barycentres (nombre choisi arbitrairement).

Comme indiqué précédemment, les vecteurs et barycentres utilisés dans cette étude sont calculés à partir de modèles construits sur le corpus français *Wikipédia 2018*. Le corpus a été préalablement lemmatisé et étiqueté syntaxiquement à l'aide du parseur *Talismane* (Urieli 2013). L'utilisation d'un corpus étiqueté nous permet de cibler uniquement les occurrences nominales de formes ambiguës (ex. *accusateur*, qui peut s'employer comme nom ou comme adjectif). Bien que cette méthode puisse induire du bruit du fait d'erreurs d'étiquetage, elle permet la prise en compte d'un nombre plus large de Nag, qui seraient sinon exclus par défaut du fait de leur ambiguïté. Par ailleurs, les modèles vectoriels étant partiellement instables (Pierrejean 2020), nous moyennons les résultats sur la base de 5 modèles construits selon les mêmes paramètres, établis par défaut (architecture CBOW, seuil de fréquence de 5, fenêtre de 5, algorithme *Negative Sampling*). Plus précisément, nous construisons un barycentre donné dans chacun des 5 modèles, puis

sélectionnons les 100 premiers voisins communs aux 5 barycentres ainsi construits.

## 2.2. Sélection des échantillons

Afin de construire les barycentres représentatifs des suffixes étudiés, nous constituons des échantillons de Nag en *-ant*, *-eur*, *-ien*, *-ier* et *-iste*. Deux conditions fondamentales sont requises pour qu'un nom figure dans un de ces échantillons. Premièrement, le Nag doit être monosémique, ou tout au moins n'avoir que des acceptions agentives. La monosémie est une condition indispensable pour éviter d'avoir des représentations distributionnelles ambiguës, qui agrégeraient en un seul et même vecteur des acceptions agentives et non agentives. Deuxièmement, notre étude se concentrant sur la construction morphosémantique de l'agentivité, nous ne sélectionnons que des Nag dont le sens agentif peut être analysé en relation avec la base morphologique. Les noms qui pourraient être lexicalement définis comme agentifs mais dont l'agentivité n'est pas construite par dérivation sont exclus. Nous considérons ici comme Nag dérivé tout nom qui dénote une entité réalisant une action en lien avec des éléments décrits par la base. Ce lien peut être de nature variée et analysé comme :

- X effectue / accomplit Base,
- X crée / modifie / détruit Base,
- X manipule / instrumentalise Base,
- X pratique / étudie Base,
- Base localise l'action de X,
- etc.

Les candidats à l'agentivité en *-ant*, *-eur*, *-ien*, *-ier* et *-iste* sont automatiquement extraits de la base de données *Lexique*<sup>9</sup>, et complétés par des données issues de la base *Lexeur*<sup>10</sup> dans le cas des dénominaux en *-eur*. Les candidats extraits sont manuellement filtrés selon une série de conditions spécifiques. Sont exclus des échantillons :

- les noms non dérivés (*peur*),
- les noms non analysables en synchronie comme étant liés à une base par une relation d'agentivité (*cordonnier*),
- les noms dont la dernière étape de construction n'est pas la suffixation<sup>11</sup> (*aide-cuisinier*),

---

<sup>9</sup><http://www.lexique.org/>

<sup>10</sup> La ressource sera très prochainement accessible en ligne à l'adresse suivante : <http://redac.univ-tlse2.fr/lexiques/>

<sup>11</sup>Nous incluons dans nos échantillons les Nag de même forme qu'un adjectif (ex. *militant*) à condition que ceux-ci puissent être analysés comme des noms suffixés, i.e. si (i) il est établi sans équivoque que leur suffixe forme effectivement des Nag (ex. *assistant*, issu d'*assister* n'est pas un adjectif), et (ii) il n'existe pas d'indication morphologique que les Nag examinés soient dérivés d'adjectifs.

- les noms en *-euse* ambigus entre féminins de *-eur* et de *-eux* (*chatouilleuse*),
- les noms dont une acception au moins n'est pas agentive, qu'ils dénotent un état (*clameur*), un habitant (*jurassien*), un membre de communauté (*mariste*), un instrument (*navigateur*), un expérimenteur (*croyant*), etc.,
- les noms multitypés ou sous-déterminés, permettant les interprétations agentives et non agentives (ex. agent / adepte, *impressionniste*).

Selon les conditions d'analyse computationnelle présentées dans la section 2.1, seuls les noms ayant au moins 5 occurrences dans le corpus *Wikipédia 2018* sont conservés. À l'issue de cette sélection, notre échantillon contient 1252 Nag utilisables comme amorces, dont 86 *-ant*, 717 *-eur*, 54 *-ien*, 161 *-ier* et 207 *-iste*<sup>12</sup>.

### 2.3. Analyse des échantillons

Avant de procéder à l'analyse distributionnelle des différents échantillons de Nag constitués, nous donnons un aperçu de leurs propriétés constructionnelles en analysant leurs bases de dérivation. Les catégories grammaticales de base<sup>13</sup> sélectionnées par chaque suffixe sont présentées dans le tableau 1. Tous les suffixes sélectionnent à la fois des bases verbales et nominales, à l'exception de *-ant* (strictement déverbal) et de *-ien* (strictement dénominal). Néanmoins, lorsque les suffixes sélectionnent les deux types de bases, ils montrent une très nette préférence pour un type : *-ier* et *-iste* pour les bases nominales, et *-eur* pour les bases verbales.

	Verbe	Nom
<i>-ant</i>	86 (100%)	-
<i>-eur</i>	681 (95,0%)	36 (5,0%)
<i>-ien</i>	-	54 (100%)

<sup>12</sup>Ces échantillons sont fortement disproportionnés. À notre connaissance, l'impact de la taille des échantillons sur l'analyse distributionnelle n'a pas été étudié, et rien ne prouve que cette disproportion affecte l'analyse. Nous choisissons ici de ne pas sous-échantillonner les données, mais plutôt de favoriser une approche maximaliste, en incluant autant d'amorces que possible, afin de garantir une bonne couverture sémantique des catégories étudiées.

<sup>13</sup>Les bases analysées sont celles qui interviennent dans la formation de Nag, indépendamment des éventuelles constructions non agentives des suffixes. Dans le cas où les Nag sont sémantiquement dérivés d'expressions polylexicales, seul un mot sert de base morphologique — généralement l'unité la plus discriminante sémantiquement (ex. *feu d'artifice* > *artificier*). Nous considérons alors que la catégorie de la base est celle de la tête de l'expression polylexicale, suivant Roché (2003) et Tribout & Amiot (2018). Ainsi, les bases en apparence adjectivales de Nag tels que *fiscaliste* et *plasticien* sont réanalysées comme des bases nominales, en tant qu'elles correspondent à *droit fiscal* et *arts plastiques*.

<i>-ier</i>	6 (3,7%)	155 (96,3%)
<i>-iste</i>	7 (3,4%)	200 (96,6%)

Tableau 1. - Catégorie grammaticale des bases dans les échantillons de Nag. Le pourcentage par suffixe est indiqué entre parenthèses.

Le sens des bases est annoté selon six types sémantiques pertinents : action (*bricoler* > *bricoleur*), objet (*machine* > *machiniste*), domaine (*histoire* > *historien*), propriété (*équilibre* > *équilibriste*), institution (*poste* > *postier*) et objet cognitif (*préface* > *préfacier*). Les bases verbales sont analysées comme dénotant des actions ou des propriétés, en fonction de leur aspect dynamique ou statif. Les bases nominales sont plus variées car elles instancient l'ensemble des types sémantiques identifiés. Leur annotation s'appuie sur un ensemble de tests linguistiques issus de la littérature (Godard & Jayez 1996, Flaux & Van de Velde 2000, Huyghe 2015, i.a.). Les noms d'action peuvent être sujet de *avoir lieu*, *se produire*, ou objet de *effectuer*, *accomplir*, *procéder à*. Les noms d'objet peuvent être sujet de *se trouver* suivi d'un locatif spatial. Les noms de domaine sont compatibles avec la construction à verbe support *faire du N*. Les noms de propriété peuvent s'employer dans *être d'un grand N*, *un état de N*, ou comme objet de *ressentir*, *éprouver*, *faire preuve de*. Les noms d'institution peuvent être sujet de *être fondé* suivi d'un locatif temporel, ou être utilisés dans des expressions comme *être nommé à la tête du N*. Les noms d'objet cognitif peuvent être objet du verbe *écrire*, ou suivis d'une relative de la forme *selon lequel P*<sup>14</sup>.

Dans le cas où les mots à analyser sont ambigus, nous annotons le sens qui se rapproche le plus de celui du nom dérivé, à condition qu'il réponde à un schéma avéré de construction de Nag. Par exemple, nous analysons *farceur* comme dérivant de *farce*<sub>1</sub> 'plaisanterie faite à quelqu'un' et non de *farce*<sub>2</sub> 'hachis d'ingrédients', sur la base de la correspondance sémantique entre les noms et de la possibilité pour des Nag dénominaux en *-eur* de se construire sur des noms d'action (ex. *bienfait* > *bienfaiteur*).

	Action	Objet	Domaine	Propriété	Institution	Obj. cog.
<i>-ant</i>	86 (100%)	-	-	-	-	-
<i>-eur</i>	686 (95,7%)	12 (1,7%)	18 (2,5%)	-	-	1 (0,1%)
<i>-ien</i>	4 (7,4%)	2 (3,7%)	46 (85,2%)	1 (1,85%)	1 (1,85%)	-

<sup>14</sup>La classe des noms d'objet cognitif inclut des noms qui ont une facette sémantique d'objet physique (ex. *roman*). Ces noms sont annotés ici comme dénotant des objets cognitifs et non des objets physiques, en raison du caractère distinctif et prépondérant de la facette cognitive.

<i>-ier</i>	21 (13%)	<b>127 (78,9%)</b>	4 (2,5%)	-	3 (1,9%)	6 (3,7%)
<i>-iste</i>	32 (15,5%)	42 (20,3%)	<b>105 (50,7%)</b>	3 (1,4%)	-	25 (12,1%)

Tableau 2. - Type sémantique des bases dans les échantillons de Nag. Le pourcentage par suffixe est indiqué entre parenthèses. Les valeurs les plus élevées par suffixe sont indiquées en gras.

Le tableau 2 présente pour chaque suffixe le type sémantique des bases sélectionnées. Il apparaît que *-ant* et *-eur* privilégient les bases actionnelles, et *-ien* et *-ier* les bases dénotant respectivement des domaines et des objets. Le suffixe *-iste* est le plus hétérogène, avec néanmoins une prédilection pour les bases qui dénotent des domaines.

### 3. ANALYSE DISTRIBUTIONNELLE

Dans cette section, nous présentons l'analyse distributionnelle des différents échantillons de Nag constitués<sup>15</sup>. Nous comparons les suffixes agentifs en analysant (i) le comportement distributionnel des différents types de Nag, et (ii) le voisinage lexical des barycentres correspondants.

#### 3.1. Densité des amorces

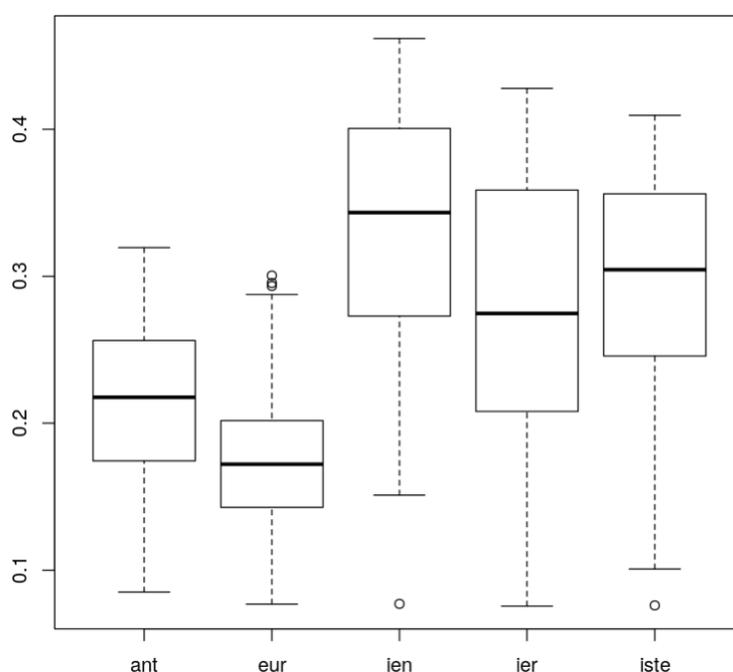
Nous commençons par examiner, pour chaque suffixe, la distribution des Nag échantillonnés selon leur proximité moyenne avec les autres Nag porteurs du même suffixe (cf. Figure 1). Cette mesure donne un aperçu de la densité des différents Nag dans l'espace vectoriel, et peut être interprétée comme un indicateur de leur cohésion sémantique. Plus les membres d'une classe de Nag sont proches dans l'espace vectoriel, plus la cohésion sémantique de cette classe est forte.

---

<sup>15</sup>Les scripts d'analyse distributionnelle, les listes de Nag utilisés dans l'étude et l'ensemble des données annotées peuvent être consultés à l'adresse suivante : <http://github.com/french-agentive-affix-rivalry>

La figure 1 indique les scores de proximité moyen minimum (trait inférieur), median (trait en gras) et maximum (trait supérieur), ainsi que les quartiles (représentés par des boîtes) pour chaque suffixe. Le score de proximité moyen figure en ordonnée. Les points représentent des valeurs aberrantes. La figure 1 indique ainsi que les noms en *-iste* ont un score de proximité moyen minimum de 0,1 (avec une valeur aberrante inférieure à 0,1), un score maximum de 0,4, et un score médian de 0,3. Cela signifie que pour 50 % des noms en *-iste*, le score de proximité avec les autres noms en *-iste* est compris entre 0,1 et 0,3, et que pour les 50 % restants, ce score est compris entre 0,3 et 0,4. La figure indique par ailleurs que 25 % des noms en *-iste* ont un score de proximité moyen avec les autres noms en *-iste* inférieur ou égal à 0,25, et 25 % un score de proximité moyen supérieur ou égal à 0,35.

Fig. 1 : Dispersion des scores de proximité moyens entre Nag du même type.



La figure 1 montre que les Nag en *-ien*, *-ier* et *-iste* sont dans l'ensemble plus proches des Nag qui partagent le même suffixe qu'eux que ne le sont les Nag en *-ant* et *-eur*. La proximité médiane varie entre 0,275 et 0,35 pour les premiers, et entre 0,175 et 0,218 pour les derniers. Ce résultat suggère que les Nag en *-ien*, *-ier* et *-iste* sont distributionnellement plus similaires entre eux que ne le sont les Nag en *-ant* et *-eur*, autrement dit, que *-ant* et *-eur* forment des Nag sémantiquement plus variés que *-ien*, *-ier* et *-iste*. On peut remarquer

que *-eur* forme la catégorie la plus hétérogène, et *-ien* la catégorie la plus homogène. Il convient toutefois de noter que les écarts entre Nag dans le cas de *-ant* et *-eur* sont moins variables que dans le cas de *-ien*, *-ier* et *-iste*. Bien qu'ils soient globalement plus proches, les Nag en *-ien*, *-ier* et *-iste* affichent de plus grandes disparités entre items. Cette particularité pourrait s'expliquer par l'existence de sous-groupes ou de séries de cas exceptionnels au sein de ces trois classes.

### 3.2. Barycentres et voisinage

L'outil distributionnel permet une analyse des Nag pris isolément, mais n'offre pas de description globale des catégories sémantiques qu'ils forment. Pour approcher une telle description, nous élaborons des représentations distributionnelles unifiées sous la forme de barycentres, comme décrit dans la section 2.1. Nous construisons un barycentre par suffixe dans chaque modèle en faisant la moyenne des vecteurs des Nag, puis nous analysons les 100 premiers voisins communs aux 5 modèles.

On retrouve parmi les 100 premiers voisins de chaque barycentre des noms qui ont servi à constituer ce barycentre, comme *postulant* pour le barycentre des Nag en *-ant*, *coiffeur* pour le barycentre des Nag en *-eur*, etc. Ce recouvrement entre amorces et voisins varie d'un suffixe à l'autre : 40%, 37% et 28% des voisins respectifs de *-ier*, *-iste* et *-eur* s'avèrent être des amorces des barycentres correspondants, contre 11% et 13% dans le cas de *-ant* et *-ien*. Si ces valeurs sont à nuancer au regard du nombre d'amorces utilisées pour chaque barycentre, elles témoignent néanmoins de la grande homogénéité morphologique de certains voisinages.

Il existe également un certain recouvrement entre les proches voisins des différents barycentres. Par exemple, *pédiatre* est un voisin partagé des barycentres de *-ien* (au 9e rang) et de *-iste* (au 6e rang), alors que *philosophe* est un voisin exclusif du barycentre de *-ien* (au 29e rang). Là encore, la portée du recouvrement varie significativement en fonction des barycentres considérés. Les barycentres de *-ien* et de *-iste*, et dans une moindre mesure de *-eur* et de *-ier*, partagent un grand nombre de voisins, avec respectivement 62 et 23 voisins communs, ce qui suggère une forte proximité entre ces barycentres. *A contrario*, les barycentres de *-ant* et de *-ien*, de *-ier* et de *-iste* ou de *-eur* et de *-iste* partagent un nombre très faible (voire nul) de voisins proches, avec respectivement 0, 1 et 3 voisins communs parmi leurs 100 premiers voisins, ce qui indique un éloignement important entre ces barycentres.

La distance entre barycentres peut être évaluée au moyen de leurs scores de proximité, présentés dans le tableau 3. Ces scores confirment la forte proximité entre *-ien* et *-iste*, et dans une moindre mesure entre *-eur* et *-ier*, ainsi que le relatif éloignement de *-ant*, notamment par rapport à *-ien* et *-iste*.

	<i>-ant</i>	<i>-eur</i>	<i>-ien</i>	<i>-ier</i>	<i>-iste</i>
<i>-ant</i>	-	<b>0,772</b>	0,534	0,656	0,502
<i>-eur</i>	0,772	-	0,745	<b>0,842</b>	0,773
<i>-ien</i>	0,534	0,745	-	0,677	<b>0,935</b>
<i>-ier</i>	0,656	<b>0,842</b>	0,677	-	0,716
<i>-iste</i>	0,502	0,773	<b>0,935</b>	0,716	-

Tableau 3. - Scores de proximité entre barycentres. Les scores les plus élevés pour chaque barycentre (en ligne) sont indiqués en gras.

En ce qui concerne les propriétés linguistiques des voisins, on peut noter d'emblée que quasiment tous les voisins analysés sont des noms<sup>16</sup>, et plus précisément des noms d'humain<sup>17</sup>. Ces noms sont en revanche très divers morphologiquement. Il peut s'agir de noms affixés (*automobiliste*), convertis (*drogué*), composés (*neuropsychiatre*), mais aussi de noms simples (*proxénète*), complexes non construits (*écrivain*) ou formés par des procédés extragrammaticaux (*indic*)<sup>18</sup>. La répartition des voisins en fonction de leur type morphologique est présentée dans le tableau 4.

Voisins de	Affixé	Convert	Composé	Simple	Complexe	Extragram.	Indét.
<i>-ant</i>	48	20	-	14	7	1	10
<i>-eur</i>	66	7	3	10	8	1	5

<sup>16</sup>Seuls deux voisins présents dans le voisinage de *-ier*, *fermier* et *chaudronnier*, ne sont pas analysés par le parseur comme des noms, mais comme des adjectifs. Il pourrait toutefois s'agir là d'une erreur d'étiquetage, notamment dans le cas de *chaudronnier*.

<sup>17</sup>À l'exception du nom *chien* présent au 76e rang du voisinage de *-eur*.

<sup>18</sup>L'analyse morphologique que nous fournissons est de type synchronique et ne prend pas en compte la formation historique des mots. Dans le cas d'une formation multiple, l'annotation proposée correspond à la dernière étape de construction. Par exemple, *couturier* (< *couture* < *coudre*) est analysé comme un nom dénominal suffixé en *-ier*. Une étiquette 'indéterminé' est utilisée pour rendre compte des mots en relation de conversion avec d'autres, dans les cas où l'orientation de la conversion ne peut pas être déterminée sur une base morphologique (ex. *assassin* / *assassiner* ou *complice* nom / *complice* adjectif, cf. Tribout 2010). Les voisins « indéterminés » sont donc soit convertis soit morphologiquement simples.

<i>-ien</i>	<b>81</b>	-	14	2	2	-	1
<i>-ier</i>	<b>80</b>	3	3	1	11	-	2
<i>-iste</i>	<b>87</b>	-	10	-	3	-	-

Tableau 4. - Type morphologique des voisins des barycentres. Les valeurs les plus élevées par suffixe sont indiquées en gras.

Le tableau 4 montre une forte présence de noms affixés parmi les voisins. Cependant, malgré une prédilection partagée pour l’affixation, tous les voisinages ne présentent pas la même diversité morphologique. Les voisins de *-iste*, *-ien*, et dans une moindre mesure *-ier* relèvent de types morphologiques moins nombreux ou plus inégalement répartis que les voisins de *-ant* et *-eur*. Certaines spécificités apparaissent. Ainsi, les noms composés sont principalement présents dans les voisinages de *-ien* et *-iste*, tandis que les noms simples se retrouvent quasi-exclusivement dans les voisinages de *-ant* et *-eur*. Les noms convertis et morphologiquement indéterminés (i.e. possiblement convertis) sont pour la plupart des voisins de *-ant*.

Les voisins affixés sont très largement des suffixés. On ne trouve dans nos données que 7 noms préfixés sur 500 voisins — la construction de l’agentivité pour les noms en question étant généralement antérieure à la préfixation (ex. *parapsychologue*). Les voisins suffixés impliquent quant à eux 11 suffixes différents (*-ain*, *-aire*, *-ant*, *-ard*, *-eur*, *-ien*, *-ier*, *-iste*, *-logue*, *-ois*, *-on*), mais dont la représentation est très inégale. Les suffixes *-ain*, *-ard*, *-ois* et *-on* sont très marginaux puisqu’ils apparaissent tous moins de 5 fois sur les 500 voisins analysés. Ces suffixes forment peu fréquemment des Nag (*forgeron*) voire ne sont pas agentifs (*villageois*, *Africain*), et leur présence dans nos données est exceptionnelle. À l’inverse, le suffixe le plus représenté parmi les voisins d’un barycentre donné est généralement celui ciblé par le barycentre en question. Ainsi, les voisins suffixés des barycentres de *-ant*, *-eur*, *-ier* et *-iste* sont formés avec le suffixe correspondant, à hauteur respectivement de 34,0%, 54,8%, 65,8% et 52,3%. Seul *-ien* ne suit pas cette tendance, qui ne se retrouve que dans 21,5% des voisins suffixés de son barycentre. On note au passage qu’il y a dans le voisinage de *-ien* et de *-iste* une forte proportion de noms en *-logue*, connus pour dénoter des spécialistes (Amiot & Dal 2007, Villoing & Namer 2012, Lasserre & Montermini 2014, Namer & Villoing 2015). Ces noms en *-logue* constituent respectivement 38,0% et 30,2% des voisins suffixés de *-ien* et de *-iste*, alors qu’ils sont totalement absents des autres voisinages analysés. On peut signaler enfin la forte présence de noms en *-iste* parmi les voisins suffixés de *-ien* (35,4%), de noms en *-ier* parmi les voisins suffixés de *-eur* (26,1%), mais aussi réciproquement de noms en *-eur* parmi les voisins suffixés de *-ier* (21,6%) et de *-ant* (25,5%).

Le tableau 5 ci-dessous indique la répartition des voisins dérivés de chaque barycentre en fonction de la catégorie grammaticale de leur base. Les voisins

dérivés sont majoritairement dénominaux et déverbaux. Les bases adjectivales sont plus rares et se retrouvent principalement dans les voisinages de *-ant* et *-eur*. La distribution entre bases verbales et nominales varie en fonction des barycentres, avec quelques tendances nettes. Les voisins de *-ien*, *-iste* et dans une moindre mesure *-ier* sont principalement dérivés de noms, à raison respectivement de 96,3%, 94,3% et 66,3% des voisins dérivés. Les voisins dérivés de *-ant* et *-eur* sélectionnent majoritairement des bases verbales, même si la tendance est moins marquée que pour les trois autres suffixes.

Voisins de	Verbe	Nom	Adjectif
<i>-ant</i>	<b>43 (63,2%)</b>	19 (27,9%)	6 (8,8%)
<i>-eur</i>	<b>38 (52,1%)</b>	30 (41,1%)	5 (6,8%)
<i>-ien</i>	3 (3,7%)	<b>78 (96,3%)</b>	-
<i>-ier</i>	26 (31,3%)	<b>55 (66,3%)</b>	2 (2,4%)
<i>-iste</i>	4 (4,6%)	<b>82 (94,3%)</b>	1 (1,1%)

Tableau 5. - Catégorie grammaticale des bases des voisins dérivés. Le pourcentage par suffixe est indiqué entre parenthèses. Les valeurs les plus élevées par suffixe sont indiquées en gras.

Le tableau 6 présente la répartition des voisins dérivés de chaque barycentre en fonction du type sémantique de leur base. Il apparaît que les bases qui dénotent des actions, des objets et des domaines prédominent. Les autres types sémantiques sont marginalement représentés, à l'exception des propriétés dans le voisinage de *-ant*. Les voisins dérivés de chaque barycentre tendent à privilégier un même type sémantique de base, en l'occurrence les actions dans le cas de *-ant* et *-eur*, les domaines dans le cas de *-ien* et *-iste*, et les objets dans le cas de *-ier*. On peut remarquer que les types sémantiques de base se répartissent de façon similaire entre les voisins de *-ien* et de *-iste* — bien que seuls ces derniers admettent les bases dénotant des objets cognitifs.

Voisins de	Action	Objet	Domaine	Propriété	Institution	Obj. cog.
<i>-ant</i>	<b>41 (60,3%)</b>	8 (11,8%)	-	13 (19,1%)	4 (5,9%)	2 (2,9%)
<i>-eur</i>	<b>45 (61,6%)</b>	16 (21,9%)	5 (6,8%)	5 (6,8%)	2 (2,7%)	-
<i>-ien</i>	4 (4,9%)	26 (32,1%)	<b>45 (55,6%)</b>	3 (3,7%)	3 (3,7%)	-

<i>-ier</i>	28 (33,7%)	<b>49 (59,0%)</b>	3 (3,6%)	2 (2,4%)	1 (1,2%)	-
<i>-iste</i>	5 (5,7%)	26 (29,9%)	<b>47 (54,0%)</b>	3 (3,4%)	2 (2,3%)	4 (4,6%)

Tableau 6. - Type sémantique des bases des voisins dérivés. Le pourcentage par suffixe est indiqué entre parenthèses. Les valeurs les plus élevées par suffixe sont indiquées en gras.

#### 4. DISCUSSION

L'analyse des propriétés morphosémantiques et des profils distributionnels des Nag échantillonnés permet d'identifier certaines spécificités des différents suffixes agentifs. Nous les passons en revue dans cette section.

##### 4.1. Proximité et homogénéité des Nag en *-ien* et *-iste*

Une forte convergence apparaît d'emblée entre *-ien* et *-iste*, ce qui confirme les observations de Dubois (1962) et Lignon (2000, 2007). Ces deux suffixes ont des profils distributionnels similaires, comme le montrent la grande proximité entre leurs barycentres, les fortes ressemblances dans la répartition des types morphologiques et des bases sémantiques de leurs proches voisins, ainsi que le nombre élevé de voisins partagés. Les bases nominales dominent parmi les amorces et les voisins dérivés, qui sont majoritairement construits à partir de noms de domaine — cette spécificité étant plus marquée encore pour *-ien* que pour *-iste*.

Les Nag en *-ien* et en *-iste* dénotent principalement des experts de domaines spécifiques et des agents d'activités intellectuelles, ce qui est confirmé par leur proximité avec les noms en *-logue*. Par ailleurs, la forte proximité entre les Nag de chaque type témoigne de leur homogénéité sémantique et, considérée conjointement avec la similarité distributionnelle entre les deux suffixes, de leur caractère distinctif relativement aux autres suffixes agentifs.

Les suffixes *-ien* et *-iste* sont donc à la fois très homogènes et très proches. Comme indiqué précédemment, *-ien* et *-iste* se différencient par des aspects morphophonologiques et diachroniques, mais on peut se demander s'ils se voient également associer des spécificités sémantiques. Un examen minutieux des amorces et des voisins de chaque suffixe fait apparaître une particularité de *-iste* concernant la dénotation d'artistes, avec de nombreuses amorces telles que (7a) et des voisins exclusifs tels que (7b).

- (7) a. artiste, violoniste, clarinettiste, marionnettiste, nouvelliste, soliste  
b. cinéaste, peintre, compositeur, dessinateur, chorégraphe, illustrateur

Cette observation corrobore un point qui transparaisait déjà dans l'étude de Cartoni *et al.* (2015). D'une manière générale, le suffixe *-iste* paraît plus diversifié que *-ien*, et il inclut plus de variations marginales (ex. *automobiliste, aubergiste, perchiste, pompiste, projectionniste*). À l'inverse, *-ien* semble plus homogène : il connaît moins de variation dans la sélection sémantique des bases et présente une densité plus forte dans l'espace vectoriel. Ses voisins exclusifs sont des noms de scientifique. On peut noter à cet égard que sur les 24 voisins exclusifs qui ne sont pas des noms en *-ien*, 10 sont des noms en *-logue*, et 2 des noms en *-graphe*.

#### **4.2. Prédilection de *-ier* pour la dénotation d'activités manuelles et commerçantes**

Le suffixe *-ier* partage un certain nombre de propriétés avec *-ien* et *-iste*, dont celles de former des noms assez similaires distributionnellement et de sélectionner préférentiellement des bases nominales. Il présente toutefois une prédilection distinctive pour les bases dénotant des objets, qui apparaît à la fois dans les amorces et les voisins du barycentre. Cette prédilection est liée à une particularité essentielle des Nag en *-ier*, qui est de dénoter très majoritairement des professions fondées sur des activités manuelles ou commerçantes. Les voisins exclusifs du barycentre de *-ier* qui ne sont pas eux-mêmes suffixés en *-ier* sont des noms tels que :

- (8) maréchal-ferrant, tanneur, aubergiste, maquignon, tailleur, tisserand, fourreur, cardeur, paveur, marchand, charron, orfèvre, droguiste, apothicaire, négociant, bûcheron, brocanteur

Comme *-ien* et *-iste*, *-ier* produit essentiellement des Nag statutaires. Cependant, l'agentivité est articulée ici autour de prédicats de fabrication, de manipulation et d'échange d'artefacts, d'où la sélection de bases dénotant ces derniers.

L'homogénéité morphosémantique des voisins et la proportion élevée d'amorces parmi les 100 premiers voisins du barycentre confirment à la fois la cohérence et la distinctivité du suffixe. On observe néanmoins une certaine proximité de *-ier* avec *-eur*, qui échappe à *-ien* et *-iste*, et qui se justifie par le fait que *-eur* peut former des noms d'ouvrier, de marchand ou d'artisan (*carreleur, ferrailleur, fraiseur, rempailleur, vendeur*). Cette proximité, que l'on relève également dans les données de Cartoni *et al.* (2015), apparaît dans la proximité des deux barycentres, et se cristallise dans des paires synonymiques (*costumier / habilleur*) ou co-hyponymiques (*cafetier / restaurateur*).

### 4.3. Polyvalence de *-eur*

Le suffixe *-eur* apparaît comme le plus hétérogène des cinq suffixes étudiés. Les Nag en *-eur* sont moins densément répartis dans l'espace vectoriel que ne le sont les autres types de Nag, ce qui témoigne de plus fortes disparités sémantiques. Les voisins proches du barycentre sont plus diversifiés morphologiquement que dans le cas de *-ien*, *-iste* et *-ier*, et comprennent notamment plus de noms simples. En outre, les catégories grammaticales de base des amorces et des voisins dérivés sont réparties de manière plus dissemblable. Par exemple, alors que les amorces en *-eur* sont majoritairement des noms déverbaux et les amorces en *-iste* des noms dénominaux, les voisins dérivés du barycentre de *-eur* ont des bases beaucoup plus variées que ceux du barycentre de *-iste*, ce qui est congruent avec une plus grande dispersion sémantique.

Une autre propriété distinctive de *-eur* est l'abondance des bases actionnelles à la fois parmi les amorces et les voisins dérivés. On peut penser que cette propriété est corrélée à une certaine polyvalence de la dénotation agentive, et donc à l'hétérogénéité sémantique. Les bases actionnelles permettent en effet d'intégrer la description d'occurrences événementielles dans la composition sémantique des dérivés. Les Nag issus de mots qui décrivent des actions peuvent ainsi dénoter des agents d'un événement particulier (agents occasionnels), et non pas uniquement des agents associés génériquement à un type d'activité (agents statutaires).

Le suffixe *-eur* apparaît donc comme le suffixe agentif le plus productif et le plus polyvalent. Il peut dénoter à la fois des agents occasionnels (*narrateur*, *expéditeur*, *vainqueur*), statutaires (*inspecteur*, *coiffeur*, *bruiteur*) et dispositionnels (*râleur*, *bagarreur*, *séducteur*). Cette polyvalence se reflète dans le voisinage du barycentre, qui inclut des Nag statutaires comme voisins partagés (*bijoutier*, *dentiste*) ou exclusifs (*coursier*, *pompiste*), des Nag occasionnels comme voisins partagés (*criminel*, *délateur*) ou exclusifs (*assassin*, *racketteur*), et des Nag dispositionnels comme voisins exclusifs (9).

- (9) escroc, farceur, ivrogne, sadique, arriviste, fêtard, voyou, drogué, arnaqueur, gaffeur, séducteur, râleur

D'après l'ensemble de nos observations, la formation de Nag dispositionnels est beaucoup plus fréquente pour *-eur* que pour les autres suffixes étudiés, et apparaît comme une spécificité sémantique de ce suffixe.

On notera pour finir qu'en dépit de sa grande polyvalence, *-eur* ne concurrence pas fortement *-ien* et *-iste*, et que la formation de noms de spécialiste reste tendanciellement distinctive de ces deux derniers.

#### 4.4. Atypie de *-ant*

Il existe un certain nombre de similarités entre *-ant* et *-eur*. Sans être aussi dispersés dans l'espace distributionnel que les Nag en *-eur*, les Nag en *-ant* sont dans l'ensemble plus distants les uns des autres que les Nag en *-iste*, *-ien* et *-ier*. Les voisins du barycentre de *-ant* sont également plus diversifiés morphosémaniquement, et comme dans le cas de *-eur*, ils comprennent plus de noms simples. Comme pour *-eur*, les bases de type actionnel dominent nettement parmi les amorces et les voisins dérivés, ce qui permet la formation régulière de Nag statutaires et de Nag occasionnels.

L'analyse distributionnelle fait néanmoins apparaître plusieurs particularités. Les noms en *-ant* sont les Nag qui sont les plus éloignés des autres dans l'espace vectoriel. D'une part, le barycentre de *-ant* est plus distant de ceux de *-ien*, *-ier* et *-iste* que ne l'est celui de *-eur*. D'autre part, ce barycentre est celui qui partage le moins de voisins proches avec les autres barycentres. En outre, il semble exister une certaine prédilection de *-ant* pour les Nag occasionnels. On remarque en effet que de nombreux Nag en *-ant*, tels que ceux sous (10a), et de nombreux voisins exclusifs de *-ant*, tels que ceux sous (10b), décrivent des agents occasionnels.

- (10) a. arrivant, assaillant, exécutant, intervenant, plaignant, poursuivant, requérant, votant
- b. émeutier, réfugié, ravisseur, client, mutin, contributeur, gréviste, agresseur

La formation de Nag occasionnels n'est certes pas l'apanage de *-ant* : *-eur* produit également des Nag occasionnels, et il existe de nombreux Nag en *-ant* statutaires. Cependant, la représentation des Nag occasionnels dans l'analyse distributionnelle des noms en *-ant* est particulièrement saillante. Cette propriété pourrait découler du fait que les Nag en *-ant* sont toujours dérivés de verbes, ce qui accroît potentiellement la proportion de Nag occasionnels formés.

On observe par ailleurs que le barycentre de *-ant* inclut dans son voisinage beaucoup plus de noms non agentifs que les autres barycentres. Des noms tels que *personne*, *villageois*, *civil*, *musulman*, *proche*, *nécessiteux*, *enfant*, *déporté*, *patient* sont des voisins exclusifs du barycentre de *-ant*. Cette particularité suggère que les Nag en *-ant* ne décrivent pas des agents prototypiques, ce que l'on pourrait éventuellement lier à la dénotation saillante d'agents occasionnels (par opposition aux statutaires). En effet, les Nag statutaires sont à la fois les plus répandus dans le lexique et les plus homogènes distributionnellement (Huyghe & Wauquier à paraître). On peut faire l'hypothèse que la présence de noms non agentifs dans le voisinage du barycentre de *-ant* est corrélée à (voire due à) la formation en nombre important de Nag occasionnels. On observe en tout cas ici une situation

paradoxe : le seul suffixe qui produit exclusivement des Nag déverbaux apparaît comme le moins prototypiquement agentif des suffixes formateurs de Nag.

## CONCLUSION

Dans cet article, nous avons étudié la construction morphologique et sémantique des noms d'agent en *-ant*, *-eur*, *-ien*, *-ier* et *-iste*, à partir de l'analyse distributionnelle d'échantillons de noms monosémiques de chaque type. Il ressort de l'étude qu'il existe des spécificités marquées pour les différents suffixes, concernant à la fois la sélection des bases et le type d'agent décrit. Les résultats confirment ou précisent certaines particularités déjà repérées dans les travaux existants (ex. la prédilection de *-ien* et *-iste* pour la formation de noms d'expert à partir de noms de domaine, la prédilection de *-ier* pour la construction de noms de métier manuel à partir de noms d'objet), et en font apparaître d'autres (ex. la polyvalence de *-eur* et sa capacité distinctive à construire des Nag dispositionnels, l'atypie de *-ant* en lien avec la formation de Nag occasionnels). D'une manière générale, il s'avère que le type d'agent décrit est étroitement lié à la sélection sémantique des bases et au mode d'élaboration de l'agentivité qui en découle.

Les spécificités identifiées sont tendancielles et n'excluent pas les cas de neutralisation. Les aléas de la construction diachronique du lexique, des processus de lexicalisation, des nécessités onomasiologiques et des interactions ponctuelles entre formes existantes (blocages lexicaux, répartition des significations, distinction des doublets, etc.) conduisent à des organisations lexicales en partie contingentes. Cet aspect est néanmoins contrebalancé par la productivité des affixes, les attractions analogiques et la néologie, qui favorisent la régularité — comme on peut l'observer par exemple dans la formation contemporaine des Nag en *-iste* et en *-ien*. Il reste que, même à un niveau de granularité sémantique moyen ou fin, on n'observe pas en synchronie de relation de un à un entre le type agentif et la forme suffixale.

Au plan méthodologique, l'approche distributionnelle que nous avons adoptée nous a permis, sur une base empirique et quantitative, de faire apparaître des tendances morphosémantiques en tant que telles. En examinant le profil distributionnel de différents types de Nag, nous avons pu à la fois évaluer l'homogénéité et la proximité des constructions sémantiques associées aux différents suffixes, et mettre en évidence les propriétés distinctives de chaque suffixe. Ce travail montre que l'association de méthodes distributionnelles et d'analyses qualitatives approfondies peut contribuer à une description sémantique fine des unités lexicales.

Dans un prochain travail, nous développerons l'analyse théorique de la concurrence morphologique entre suffixes agentifs. Nous étudierons la portée exacte de la corrélation entre les propriétés sémantiques et morphologiques

des Nag, notamment en employant des méthodes inductives, fondées sur le clustering distributionnel.

#### **BIBLIOGRAPHIE**

- AMIOT D., DAL G. (2007). Integrating Combining Forms into a Lexeme-Based Morphology. In : G. Booij, B. Fradin, A. Ralli & S. Scalise (eds), *Online Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*, 323-336.
- ANDREWS S. (1996). *The Semantics of Suffixation. Agentive substantival Suffixes in Contemporary Standard Russian*. München/Newcastle: LINCOM EUROPA.
- ANSCOMBRE J.-C. (2001). A propos des mécanismes sémantiques de formation de certains noms d'agent en français et en espagnol. *Langages*, 143, 28-48.
- ANSCOMBRE J.-C. (2003). L'agent ne fait pas le bonheur : agentivité et aspectualité dans certains noms d'agent en français et en espagnol. *Thélème, Revista Complutense de Estudios Franceses*, 11-27.
- BAUER L., LIEBER R., PLAG I. (2013). *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- BENVENISTE E. (1948). *Noms d'agent et noms d'action en indo-européen*. Paris : Maisonneuve.
- BOLEDA G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6, 213-234.
- BOOIJ G. (1986). Form and meaning in morphology: the case of Dutch agent nouns. *Linguistics*, 24, 503-517.
- BOOIJ G. (2002). *The Morphology of Dutch*. Oxford: Oxford University Press.
- CARTONI B., NAMER F. (2012). Linguistique contrastive et morphologie : les noms en *-iste* dans une approche onomasiologique. In : F. Neveu, V. Muni Toke, P. Blumenthal, T. Klingler, P. Ligas, S. Prévost, S. Teston-Bonnard (éds), *Actes du CMLF 2012*, 1245-1260.
- CARTONI B., NAMER F., LIGNON S. (2015). A cross-linguistic insight on agentive noun formation in Italian and French. *Selected Papers from the 8th Décembrettes: Morphology in Bordeaux, Carnets de Grammaire*, 22, 81-98.
- CORBIN, D. (1987). *Morphologie dérivationnelle et structuration du lexique (Vol. Tomes 1 & 2)*. Tübingen : Max Niemeyer Verlag
- CORBIN D., CORBIN P. (1991). Un traitement unifié du suffixe *-ier(e)*. *Lexique*, (10), 61-145.
- CRUSE D.A. (1973). Some thoughts on agentivity. *Journal of Linguistics*, 9, 11-23.
- DELANCEY S. (1984). Notes on agentivity and causation. *Studies in Language*, 82, 181-213.

- DOWTY D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547-619.
- DUBOIS J. (1962). *Étude sur la dérivation suffixale en français moderne et contemporain: essais d'interprétation des mouvements observés dans le domaine de la morphologie des mots construits*. Paris: Larousse.
- FILLMORE C. (1968). The case for case. In: Bach E, Harms RT (eds), *Universals in linguistic theory*. New-York: Holt, Rinehart & Winston, 1-88.
- FIRTH J.R. (1957). A synopsis of linguistic theory, 1930-1955. In: J.R. Firth (ed.), *Studies in Linguistic Analysis*. Oxford: Basil Blackwell, 1-32.
- FLAUX N., VAN DE VELDE D. (2000). *Les noms en français: esquisse de classement*. Paris: Editions Ophrys.
- FRADIN B., KERLEROUX F. (2003). Troubles with lexemes. In : G. Booij, J. Cesaris, S. Scalise & A. Ralli (eds), *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting*. IULA-Universitat Pompeu Fabra, 177-196.
- GODARD D., JAYEZ J. (1996). Types nominaux et anaphores : le cas des objets et des événements. In : W. De Mulder, L. Tasmowski-De Ryck & C. Vetter (éds), *Anaphores temporelles et (in-)coherence, Cahiers Chronos 1*, Amsterdam: Rodopi, 41-58.
- GRUBER J.S. (1967). Look and see. *Linguistics*, 43, 937-947.
- HARRIS Z.S. (1954). Distributional structure. *Word*, 10(2-3), 146-162.
- HUYGHE R. (2015). Les typologies nominales : présentation. *Langue Française*, 185, 5-27.
- HUYGHE R. (à paraître). La construction morphosémantique des noms d'agent en français. In: A. Aleksandrova & J. P. Meyer (éds), *Nommer l'humain : descriptions et catégorisations*. Paris: L' Harmatthan.
- HUYGHE R., TRIBOUT D. (2015). Noms d'agents et noms d'instruments : le cas des déverbaux en -eur. *Langue Française*, 185, 99-112.
- HUYGHE R., WAUQUIER M. (à paraître). What's in an agent? A distributional semantics approach to agent nouns in French. *Morphology*. DOI: 10.1007/s11525-020-09366-2
- KERLEROUX F. (2004). Sur quels objets portent les opérations morphologiques de construction ? *Lexique*, 16, 85-123.
- KIPPER SCHULER K. (2005). *VerbNet. a broad-coverage, comprehensive verb lexicon*. (Doctoral dissertation, University of Pennsylvania)
- KNITTEL M.L. (2017). *French derived nominals in -ant: semantic properties*. Talk given at ISMO, Lille, December 2017.
- LASSERRE M., MONTERMINI F. (2014). Pour une typologie des lexèmes construits : entre composition, composition néoclassique et affixation. F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefer, S. Prévost (éds), *Actes du CMLF 2014*, 1797-1812.

- LENCI A. (2018). Distributional models of word meaning. *Annual review of Linguistics*, 4, 151-171.
- LIGNON S. (2000). La suffixation en *-ien* : aspects sémantiques et phonologiques. (Doctoral dissertation, University Toulouse Jean Jaurès)
- LIGNON S. (2007). Les noms de spécialistes en *-iste* et en *-ien* : le chimiste perturbé ou comment le physicien se réajuste. In : B. Vaxélaire, R. Sock, G. Kleiber & F. Marsac (éds), *Perturbations et Réajustements. Langue et langage*. University Marc Bloch - Strasbourg 2, 287-295.
- LIGNON S., NAMER F. (2018). Catégories d'humains entre dictionnaire et usage : le *wagnérien* est-il un partisan ou un spécialiste ? In : C. Schnedecker, W. Mihatsch (éds), *Les noms d'humains : théorie, méthodologie, classification*. Berlin : De Gruyter, 278-300.
- LUSCHÜTZKY H.C. (2011). Agent-noun polysemy in Slavic: some examples. *STUF-Language Typology and Universals Sprachtypologie und Universalienforschung*, 64(1), 75-97.
- MIKOLOV T., CHEN K., CORRADO G., DEAN J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:13013781*
- MIKOLOV T., YIH W., ZWEIG G. (2013b). Linguistic regularities in continuous space word representations. In : *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 13, 746-751.
- NAMER F., VILLOING F. (2015). Sens morphologiquement construit et procédés concurrents : les noms de spécialistes en *-logue* et *-logiste*, *Revue de Sémantique et Pragmatique*, 35-36, 7-26.
- PIERREJEAN B (2020). *Qualitative Evaluation of Word Embeddings: Investigating the Instability in Neural-Based Models*. (Doctoral dissertation, University Toulouse Jean Jaurès)
- ROCHÉ M. (1997). *Briard, bougeoir et camionneur* : dérivés aberrants, dérivés possibles. In : D. Corbin, B. Fradin, B. Habert, F. Kerleroux & M. Plénat (éds), *Mots possibles et mots existants. Ires rencontres du forum de morphologie*, 241-250.
- ROCHÉ M. (2003). Catégorisation et recatégorisation en morphologie dérivationnelle: le cas de la dérivation en *-ier(e)*. In : G. Coll & J.P. Régis (dir.), *Morphosyntaxe du lexique. Catégorisation et mise en discours, Actes du Colloque de Tours, 7-8 juin 2002, Travaux Linguistique du CerLiCO*, 16, 75-92.
- ROCHÉ M. (2004). Mot construit ? mot non construit ? quelques réflexions à partir des dérivés en *-ier (e)*. *Verbum*, 26(2), 459-480
- ROCHÉ M. (2011). Quel traitement unifié pour les dérivations en *-isme* et en *-iste* ? In : M. Roché, G. Boyé, N. Hathout, S. Lignon, & M. Plénat (éds), *Des unités morphologiques au lexique*. Paris: Hermès, 69-143.

- ROSENBERG M. (2008). *La formation agentive en français. les composés [VN/A/Adv/P]N/A et les dérivés V-ant, V-eur et V-oir (e)*. (Doctoral dissertation, University of Stockholm)
- ROY I., SOARE E. (2012). L'enquêteur, le surveillant et le détenu : les noms déverbaux de participants aux événements, lectures événementielles et structure argumentale. *Lexique*, 20, 207-231.
- SCHLESINGER I.M. (1989). Instruments as agents: on the nature of semantic relations. *Journal of Linguistics*, 25(1), 189-210.
- TRIBOUT D. (2010). *Les conversions de nom à verbe et de verbe à nom en français*. (Doctoral dissertation, University of Paris Diderot)
- TRIBOUT D., AMIOT D. (2018). Nom et/ou adjectif? Quelle catégorie d'output pour les suffixés en-iste? In : O. Bonami, G. Boyé, G. Dal, H. Giraud, & F. Namer (eds), *The lexeme in descriptive and theoretical morphology*. Berlin : Language Science Press, 69.
- ULLAND H. (1993). *Les nominalisations agentive et instrumentale en français moderne*. Berne : Peter Lang.
- URIELI A. (2013). *Robust French syntax analysis: reconciling statistical methods and linguistic knowledge in the Talisman toolkit*. (Doctoral dissertation, University Toulouse Jean Jaurès)
- VAN VALIN R., LAPOLLA R. (1997). *Syntax. Structure, Meaning and Function*. Cambridge : Cambridge University Press.
- VILLOING F., NAMER F. (2012). Composition néoclassique en *-logue* et *-logiste* : les noms en *-logue* sont-ils encore des noms de spécialistes. *Verbum*, 34(2), 213-231.
- WAUQUIER M., HATHOUT N., FABRE C. (2020). Contributions of distributional semantics to the semantic study of French morphologically derived agent nouns. In : J. Audring, N. Koutsoukos, & C. Manouilidou (Eds.), *Rules, patterns, schemas and analogy, MMM12 online proceedings, vol. 12*, 111–121.
- WINTHER A. (1975). Note sur les formations déverbaux en *-eur* et *-ant*. *Cahiers de lexicologie*, 26, 35-54.