



What's in an agent?

A distributional semantics approach to agent nouns in French

Richard Huyghe¹ · Marine Wauquier²

© The Author(s) 2020

Abstract This article investigates the morphological diversity of agent nouns (ANs) in French. It addresses the questions of which nouns form a semantically coherent class of ANs, what their morphological properties are, and whether these properties correlate with agentive subtypes. To deal with these issues, a distributional semantics approach is adopted. The investigation is based on the distributional study of monosemous deverbal ANs ending in *-eur*, and on the examination of word similarities in the French Wikipedia corpus. It is shown that French ANs as a homogeneous distributional class display a large variety of morphological profiles. ANs can be affixed, converted, compound nouns, as well as opaque and morphologically simple nouns. Agentive affixes are diverse and correlated to the selection of bases from different lexical categories and semantic types. It appears that agent meaning in the nominal domain is not necessarily imported from the verbal domain, but can develop directly in the semantic structure of nouns. In addition, a distinction between functional, occasional and behavioral ANs, depending on whether they denote agents with an occupational status, agents in a particular event, or agents with a tendency to act in a certain way, is proven to be distributionally relevant. This distinction applies to all ANs, possibly in correlation with specificities as regards morphological type, base and affix selection. The study illustrates that with a careful processing of linguistic data, distributional semantics can help answering basic research questions and support fine-grained theoretical distinctions.

Keywords Agent noun · Distributional semantics · French · Agentive suffix

✉ R. Huyghe
richard.huyghe@unifr.ch

M. Wauquier
marine.wauquier@univ-tlse2.fr

¹ University of Fribourg, Fribourg, Switzerland

² CLLE, CNRS & University of Toulouse Jean Jaurès, Toulouse, France

1 Introduction

The class of agent nouns, i.e. nouns that describe performers of actions, is not clearly delimited in the existing literature. Its definition raises a number of issues that call for in-depth examination, and the extension of the class is rarely discussed as such. In morphological studies, the category of agent noun (henceforth AN) has been used mostly to characterize the lexical output of some derivational processes. However, whether ANs necessarily result from morphological processes, and what these processes can possibly be deserve further investigation. In many languages, some representative instances of ANs are identified, that are usually deverbal nouns ending with a given suffix. The existence of a prototypical agentive suffix¹ is generally admitted for languages such as Latin (*-tor*), English (*-er*), Spanish (*-dor*), Dutch (*-er*), German (*-er*), etc. Yet conceptions vary as to whether the class of ANs extends to other morphological constructions. Lists of agentive suffixes are sometimes given for one language or another, but they differ according to the authors, and the criteria for identifying ANs are rarely made explicit—one may question, for instance, the lexical category of the base, since the suffixes presented as agentive (e.g. *-er*, *-ist*, *-an*, *-eer* in English) are often not strictly deverbal devices.

As far as French is concerned, ANs have been claimed to be deverbal nouns ending in *-eur*, as well as verb-noun compounds, or nouns ending in *-aire*, *-ien*, *-ier*, *-iste*, be they deverbal or not (Dubois 1962; Winther 1975; Anscombe 2001; Roché 2003, 2011; Villoing 2009; Roy and Soare 2014; Cartoni et al. 2015; Schnedecker and Aleksandrova 2016 to mention a few). It has also been suggested that underived nouns such as *avocat* ‘lawyer’ or *médecin* ‘doctor’ should be considered ANs (Huyghe to appear). Most existing studies focus on one or two suffixes and do not discuss the whole range of morphological constructions that yield ANs, nor do they question the use of the ‘agent’ label in a variety of cases. The agentivity of some nouns is nonetheless overtly discussed, as in the case of nouns ending in *-ant* (e.g. *manifestant* ‘demonstrator’, *combattant* ‘fighter’). Authors like Anscombe (2003) and Roy and Soare (2012) argue that *-ant* deverbal nouns are not agentive, whereas others, like Rosenberg (2008), take the opposite view. The delineation of the AN category is rendered difficult by the absence of consensual definition and linguistic tests to formally identify ANs. Furthermore, it is largely acknowledged that prototypical ANs divide into distinct semantic subclasses, which questions both the consistency of the class and the possible correlation between the morphological characterization and the semantic subclassification of ANs.

In this article, we investigate the morphological diversity of French ANs and the relation between form and meaning as concerns agent denotation. We aim to determine which nouns form a semantically coherent class of ANs, and to what extent AN classification depends on morphological construction, especially on affix and base selection. We also address the question of AN subclassification, and explore the correlation between the morphological and semantic properties of AN subclasses.

To deal with these issues, we adopt a distributional semantics approach, and make use of machine learning tools developed in computational linguistics. We investigate

¹The term *agentive suffix* is used for writing convenience, to denote suffixes that are used to form ANs. We do not take a position here in the long-standing debate about the semantic content of affixes.

whether nominal agentivity can be highlighted by distributional analysis, and whether distributional clues can support a semantic typology of ANs. Our general methodology will be based on the morphosemantic analysis of prototypical *-eur* ANs, and of the words that are the most similar to them in distributional semantics models. It will be shown that (i) French ANs as a homogeneous distributional class display a large variety of morphological profiles, including derived and underived nouns, (ii) agentive affixes are multiple and correlated to the selection of bases from different lexical categories and semantic types, (iii) a three-way partition of ANs, depending on whether they denote functional, occasional or behavioral agents, is both semantically and morphologically relevant.

The article is organized as follows. In Sect. 2, we examine theoretical issues about agentivity, and the definition and subclassification of ANs. Section 3 presents the methodology we use in the distributional semantics investigation. In Sect. 4, we describe the extension of the class of ANs, basing on distributional similarities with prototypical monosemous ANs. In Sect. 5, we assess the distributional relevance of AN subclasses and their correlation with the morphosemantic properties of nouns.

2 Defining agent nouns

Identifying agent nouns in a given language is not a trivial task. First, it requires some discussion of the notion of agent, which has been the subject of considerable controversy among syntacticians and typologists. Second, agent as a semantic role is designed to account for verb-argument relations, and its application to the lexical analysis of nouns is not straightforward and needs evaluation. Third, the semantic consistency of the AN category is challenged by the existence of distinct homogeneous groups of ANs. These three issues are successively tackled in this section.

2.1 Agentivity in question

Agent is reputedly a difficult notion to define. Some major properties of agents have been long-debated, and different criteria have been discussed in the literature to characterize agents (animacy, control, volition, instigation, sentience, performance, accountability, motion, etc.). However, there is no consensus neither on the relevant criteria, nor on the values these criteria should be assigned to accurately define agents. Authors notably disagree on the animacy and intentionality of agents. In Gruber's terms, an agentive verb is a verb "whose subject refers to an animate object which is thought of as the willful source or agent of the activity described by the verb" (1967:943). But animacy has been contested e.g. by Cruse (1973:11) who claims that "comparing, for instance, *John overturned the dustbin* and *The wind overturned the dustbin*, it is difficult to see how *the wind* is less an agent than *John*"; and authors like Schlesinger (1989:194) have argued against intentionality, alleging that "it is generally recognized that intention is not a necessary feature of agentivity, for after all we often do things accidentally".

Another difficulty arises from the fact that semantic roles are used both in contextual analyses and in verbal descriptions. Verbs selecting subjects that can be animate and inanimate, intentional and unintentional, instigator and non-instigator, etc.

should be assigned one or several lexical entries depending on the defining conditions of agentivity, even though their core meaning (e.g. causative) remains unchanged. Such considerations lead Holisky (1987) and Van Valin and Wilkins (1996) to argue that agentivity is not determined by the semantic structure of verbs, but is in most cases pragmatically established. Still many lexical descriptions define verbs as (non-)agentive, and in lexical resources such as Verbnets (Kipper Schuler 2005) or PropBank (Palmer et al. 2005), verbs are characterized with regard to agent role assignment.

A way to deal with these definition issues is to develop a prototypical model of agentivity, as suggested by Lakoff (1977), DeLancey (1984), Dowty (1991), among others. The idea that the definition of agent needs to be prototypically structured can be found throughout the literature on semantic roles, from Fillmore (1968:46) claiming that agents are “typically” animate instigators, to Kipper Schuler (2005:31) describing agents as “generally” human or animate, and “mostly” volitional. Prototypes are also compatible with a scalar approach to agentivity, in which degrees of agentivity are distinguished according to the satisfaction of different defining conditions, and some arguments or participants are analyzed as more agentive (i.e. more prototypically agentive) than others (Grimm 2011).

It is not our purpose here to extensively discuss the criteria for agentivity, but only to identify some definition issues that could blur the lexical use of the notion. To proceed with an operational notion of agent and not be a priori restrictive about the lexical items considered agentive, we will base our study on a broad conception of agent. We roughly define here agents as effectors (i.e. entities deploying energy to perform actions) that are prototypically, but not necessarily, animate and intentional. A definition of ‘effector’ is given by Van Valin and LaPolla (1997:118) as “the participant that brings something about, but there is no implication of its being volitional or the original instigator”. Assimilating the core notion of agent to that of effector is in line with Cruse’s definition of the ‘agentive’ feature (as opposed to ‘volitive’, ‘effective’ and ‘initiative’) as “referring to an action performed by an object which is regarded as using its own energy in carrying out the action” (1973:21).

The definition given here is intended to be suitable for both contextual and lexical analyses (see Sect. 2.2). It preserves the widespread view that many causative verbs can be described as agentive, even though their subject is not necessarily animate and volitional. More broadly, it satisfies the condition of defining semantic roles independently of ontological features, so that role assignment and selectional restrictions can be clearly distinguished in predicate analysis.

Note that, although adopting a prototypical structure, the definition we propose is not similar to the one given by Dowty (1991), which is based on proto-roles. Dowty includes sentience or perception as a defining feature of ‘proto-agents’, thus analyzing subjects in *John knows/believes/is disappointed at the statement* and *John sees/fears Mary* as proto-agentive. Experiencers will not be considered agents in the more fine-grained perspective we adopt here. We assume indeed that agents are involved in dynamic situations. The property of dynamicity vs. stativity appears to be more firmly established in the verbal lexicon than intentionality or even animacy—in French it can be reliably tested through the construction with the progressive. Hence subjects such as *these columns* in *These columns support the weight of the pediment*

(Cruse 1973:19) will not be considered agents. The condition on dynamicity also crucially distinguishes agents from causes, since causes can be stative. According to a generally shared view, agentivity implies causation, but not reciprocally, which is congruent with the conception of agent advocated above.

Dismissing animacy as a necessary condition could be an issue when distinguishing agents from instruments, since agents and instruments are often defined respectively as animate and inanimate entities. The classification of inanimate entities that are on the edge between instrumentality and agentivity (e.g. sophisticated mechanical devices) may be problematic. We will maintain a straight distinction by defining instruments as entities that are fundamentally used by other entities to perform actions.² Furthermore, including a necessary animate condition in the definition of agents would create a blind spot for natural forces and for biological, chemical or abstract agents that carry out actions autonomously. A complementary role would be necessary to account for such agents,³ and again the analysis would have to distinguish between two types of role assignment in case a predicate is compatible with both animate and inanimate effector subjects.

2.2 From agent to agent nouns

Semantic roles are fundamentally intended to capture verb-argument relations. They can be used in the analysis of verbal properties and contextual relations. For instance, *my sister* and *the housebreaker* in (1) denote respectively the agent and the patient of the verb *knock out*.

(1) *My sister knocked out the housebreaker.*

The use of semantic roles in the lexical analysis of nouns implies both abstraction from contextual relations and a shift from verbal to nominal properties. At the lexical level, nouns can be classified as ANs if they describe their referent as the (potential or actual) agent of an intrinsically specified action. In that perspective, *sister* will not be classified as AN, but rather as a relational noun (Barker 2008), as opposed to *housebreaker*, which is per se an AN, not a patient noun.

The classification as AN requires two conditions to be satisfied: the presence of an action component in the semantic structure of the noun, and the description of the corresponding agent. These conditions can be difficult to assess. A way to guarantee

²It is known that under some conditions instrument-denoting nouns can be used in subject position (Fillmore 1968; Schlesinger 1989; Kamp and Rossdeutscher 1994; Grimm 2013, among others), in which case they are arguably assigned the role of cause or agent (Alexiadou and Schäfer 2006; Ježek and Varvara 2015). It is crucial to note that deverbal instrument nouns cannot be essentially defined in correspondence with the subject of their base verbs. For instance, French noun *perceuse* 'drill' fundamentally relates to Z, not to X, in *X perce Y (avec Z)* 'X drills Y (with Z)'. Consequently, instrument nouns as such do not verify the 'external argument generalization' hypothesis that claims that suffixes like English *-er* nominalize the external argument of the base verb. In a psycholinguistic study, Lowder and Gordon (2015) have shown that instruments in subject position are more unexpected and need more cognitive processing than natural forces and animate agents.

³This is to some extent the case for the 'originator' role used by Lieber and Andreou (2018:195) to classify "internal causers and non-animate external causers, saving 'agent' for animate external causers with a degree of sentience and volitionality". *-Er* nominals in sentences like *Class, as always, is the great unifier, and great divider, of British society* are then analyzed as originators.

them is to assume that ANs are deverbal nouns that denote the agentive argument of the base verb. Such an option offers the advantage of a controlled definition, since it strictly limits ANs to deverbal nouns. However it introduces a major difficulty in that it divides morphologically heterogeneous but semantically homogeneous classes:

- (2) a. *sculpteur* ‘sculptor’, *guérisseur* ‘healer’, *rédacteur* ‘writer’, *manifestant* ‘demonstrator’, *protestataire* ‘protestor’
 b. *artiste* ‘artist’, *médecin* ‘doctor’, *scribe* ‘scribe’, *gréviste* ‘striker’, *émeutier* ‘rioter’

Nouns in (2a) being derived from dynamic verbs and denoting their agents, they will be considered ANs, but their non-deverbal quasi-synonyms or closely related nouns in (2b) will not. Such an option seems inconsistent if one aims to define a semantically coherent class of ANs.

On the other hand, dismissing the deverbal condition raises the issue of control over the class of ANs. If nouns as morphologically diverse as those in (2b) and (3) are to be considered ANs, how can we ascertain their agentive meaning and consequently delineate the class of ANs?

- (3) *historien* ‘historian’, *bijoutier* ‘jeweler’, *juriste* ‘legal expert’, *prophète* ‘prophet’, *tribun* ‘orator’, *disquaire* ‘record dealer’, *tyran* ‘tyrant’, *cardiologue* ‘cardiologist’

To our knowledge, there is no reliable linguistic test proposed in the literature to clearly identify ANs. ANs should be defined as denoting the agent of a dynamic predicate, but the evaluation of such a property could be arbitrary. How can we decide whether nouns such as *chef* ‘boss’, *gourou* ‘guru’, *milicien* ‘militiaman’, *dandy* ‘dandy’, *salarié* ‘employee’, *moine* ‘monk’, *chenapan* ‘rascal’, *héros* ‘hero’ should be defined with reference to agentivity? A corollary question concerns how the action component is generated in the semantic structure of ANs, in case it is not directly inherited from the base. It could be assumed that some derived ANs involve a general dynamic predicate that is determined by the semantics of the base. For example, artifact-denoting bases would trigger predicates of manufacturing, handling or providing, as in *chocolatier* ‘chocolate maker’, *bagagiste* ‘luggage handler’, *libraire* ‘bookseller’. But in the case of morphologically simple ANs, the action component has to be created in the semantic structure of the noun without any external input, and a theory of nominal agentivity should account for such a possibility (Huyghe to appear).

2.3 Subclasses of agent nouns

The semantic consistency of the class of ANs can be addressed with respect to its extension, but also to the differences that exist between its members. It is often acknowledged that prototypical ANs divide into distinct groups, which raises further questions about a possible correlation between morphological and semantic subclasses of ANs.

Benveniste (1948) comments on the existence of two morphological types of ANs in Indo-European, basing on data from Vedic, Avestan and ancient Greek. According to him, these two types correlate with a pervasive semantic distinction between

two kinds of agents: ‘the doer of an act’ as opposed to ‘the agent of a function’. The former is defined as the agent in a particular event, while the latter is assigned a functional status, regardless of the actual performing of an action. Although the morphological distinction is not observed in Latin, Benveniste argues that the semantic distinction remains, and can be detected in contemporary languages like French—possibly in connection with base allomorphy, as in *sauveur* ‘savior’ (doer of an act) vs. *sauveteur* ‘rescue worker’ (agent of a function).

A somewhat similar distinction can be found in the literature about English *-er* nominalizations. Rappaport Hovav and Levin (1992) distinguish two major subclasses of *-er* nominals, i.e. eventive and non-eventive nominals, depending on their referring to an actual event or not. Agentive nominalizations divide into the two subclasses, whether they denote the actual agent in a specific event or not, as in *saver of lives* vs. *lifesaver*. Rappaport Hovav and Levin claim that the distinction between eventive and non-eventive nominals correlates with the (non-)availability of complement structure: only eventive nominals have complements that denote participants in the underlying event. That syntactic analysis is discussed by Alexiadou and Schäfer (2010), who argue that both types of nominals have the same rich syntactic structure, and that the difference lies in aspectual specification. They suggest that eventive and non-eventive *-er* nominalizations actually differ with respect to episodic vs. dispositional aspect. Roy and Soare (2012) propose a similar analysis for *-eur* nominalizations in French—with the difference that instrumental *-eur* nouns are not considered to be dispositional. According to them, the difference between episodic and dispositional *-eur* ANs depends on the specific vs. generic interpretation of the (sometimes unexpressed) complement, as in *le vendeur de ce bien immobilier* ‘the seller of this real estate’ vs. *un vendeur de journaux* ‘a newspaper seller’.

The distinction between episodic and dispositional agents allows for a syntactic account of nominalization constructions. It is nevertheless possible to draw a lexical distinction between ANs that denote agents involved in a specific event (which in order to avoid confusion with syntactic analysis we will call ‘occasional agents’) and ANs that denote agents with a functional status (which we will call ‘functional agents’).⁴ Occasional ANs can be used with a specific complement structure as in (4), whereas functional ANs in French can be used as bare predicates (Van Peteghem 1993; Matushansky and Spector 2005; De Swart et al. 2007; Castella 2014) as in (5).

- (4) *l'agresseur de la vieille dame, l'expéditeur de ce colis, le fondateur de cette ville*
 ‘the old lady’s attacker’, ‘the sender of this parcel’, ‘the founder of this city’
- (5) *Pierre est {inspecteur / déménageur / tatoueur}*.
 ‘Pierre is {an inspector/a mover/a tattooist}’ (lit. ‘Pierre is inspector’)

A third class of ANs has been introduced by Huyghe and Tribout (2015) to account for ANs such as *séducteur* ‘seducer’, *bagarreur* ‘fighter’, *bosseur* ‘hard worker’,

⁴In the examples provided by Roy and Soare (2012), the same nouns are used in competing structures, and it seems that the difference between episodic and dispositional agent nominals pertains to context. We assume on the contrary that some semantic properties of *-eur* deverbal nouns are established in the lexicon, thus determining their tendencies of use and interpretation. These tendencies can be quantified, and one of the goals of the present paper is precisely to highlight them through distributional analysis.

gaffeur ‘blunderer’, *bluffeur* ‘bluffer’, that denote agents with a propensity to do certain things or to act in a certain way. Nouns of this type are compatible with adjectives in a non-intersective (i.e. adverbial) interpretation, as illustrated in (6).

- (6) *un grand séducteur, un gros buveur, un énorme râleur*
 ‘a great seducer’, ‘a heavy drinker’, ‘a huge complainer’

Size adjectives in (6) characterize the performing of actions, especially as regards frequency, and they can have a habitual reading (e.g. *un grand séducteur* ‘a great seducer’ refers to somebody who seduces a lot). These nouns denote what we will call ‘behavioral agents’. They can easily be converted into adjectives, and somehow fall in between occasional and functional ANs.

It can be asked whether the distinction between functional, occasional and behavioral ANs can be observed through large corpus data, and whether it applies to non-*eur*-suffixed ANs. To answer questions as to (i) which nouns exactly should be considered ANs, (ii) which semantic subclasses of ANs are relevant and whether these subclasses correlate with morphological properties of ANs, we will adopt a distributional semantics approach, and compare the linguistic behavior of different AN candidates in a corpus-driven study. Two investigations will be carried out with respect to the morphological diversity of ANs (Sect. 4) and to AN subclassification (Sect. 5). Beforehand, we briefly present our research methodology.

3 Methodology

Following the distributional hypothesis, we assume that the agentive meaning of ANs determines their general distribution (i.e. not only in the syntactic position of agent), and conversely that the distributional profile of a noun reveals its possible agentive meaning. Building on that assumption, we develop a method to approach AN meaning that is based on the distributional representation of prototypical ANs. We first present theoretical prerequisites for distributional analysis (Sect. 3.1). Then we describe the experimental set-up implemented in the two investigations that will be conducted (Sect. 3.2).

3.1 Theoretical prerequisites

Distributional semantics provides a geometric representation of word meaning based on the statistical distribution of words in corpus. It is grounded in the distributional hypothesis which states that the meaning of a word can be assessed through its distribution. A difference in distribution between two words is said to correlate with a difference in meaning (Harris 1954; Firth 1957), and it is assumed accordingly that the semantic similarity of words can be estimated by the amount of context they share.

According to their co-occurrences in corpus, words are represented as vectors in a multidimensional space (also called ‘vector space’ or ‘distributional semantics model’). The distributional and by extension semantic proximity between two words

is computed as the cosine distance between their vectors, ranging from 0 for no proximity to 1 for maximum proximity (i.e. the two vectors being identical). This proximity score allows for the ordering of the nearest neighbors of a given vector, i.e. the words that are supposed to be the most similar. Because of their mathematical nature, word vectors can moreover be combined so as to approximate compositionality, and undergo algebraic operations such as addition, subtraction or averaging (Lenci 2018; Boleda 2020). In recent years, distributional semantics models underwent a renewal with the development of neural networks-based tools such as Word2Vec (Mikolov et al. 2013a). While these new methods imply a loss in model comprehension and the need for quantitatively larger corpora, they proved to be computationally cheaper and just as efficient, if not more, in many natural language processing tasks.

Distributional semantics has gradually established itself as a major tool to study various linguistic phenomena on a large scale. For instance, Erk et al. (2010) and Lenci (2011) deal with the issue of verb argument thematic fit, at the interface of syntax and semantics. They assess the thematic fit of a target noun as an argument of a given verb, basing on its similarity to a set of attested arguments in corpus. Distributional semantics has also been used to study the interaction between morphology and semantics, in which case works engage in the semantic analysis of morphologically implemented classes. Many studies aim at semantically distinguishing rival word-formation patterns (Varvara et al. 2016; Wauquier et al. 2020), while others address affix polysemy and word sense disambiguation (Lapesa et al. 2018). Such studies benefit from the automatic corpus-driven analysis provided by distributional semantics, and carried out over large sets of data. The level of abstraction from individual forms and occurrences that is inherent to distributional semantics allows for generalizations over entire classes, and yet sheds light on more local, word-specific phenomena.

3.2 Experimental set-up

In this work, we use distributional semantics to outline the semantic behavior of ANs as a class. We do not approach ANs individually, but on a larger scale, basing on the assumption that the agentive meaning shared by ANs is accessible through their general distribution. Following Kintsch (2001), Erk et al. (2010), Wauquier et al. (2018), Bonami and Paperno (2018), Mickus et al. (2019), among others, we hypothesize that a unified representation of the AN class can be computed by averaging the vectors of its members. The average vector (or centroid) can be seen as representative of the semantics of ANs, because it aggregates their distributional and by extension semantic properties. The analysis of the centroid's nearest neighbors provides insight on the semantic properties it conveys, and therefore presumably on those of the targeted class.

The models we use are trained on the French Wikipedia corpus (dump from October 1, 2018). Containing about 900 million words, it is one of the largest French corpora, and it displays a broad variety of subjects and vocabulary. This diversity ensures a large lexical coverage. The corpus was beforehand lemmatized and part-of-speech (henceforth POS) tagged with Talismane tagger (Urieli 2013). Because it

relies on supervised machine learning, this pre-processing comes with lemmatization and POS tagging errors. The system achieves an accuracy in POS tagging ranging from 93.63% to 97.55% depending on the configuration and the corpus (Urieli 2013:141). While we acknowledge these limitations, the use of an automatically annotated corpus allows for syntactic disambiguation of target words in distributional analysis, as detailed in Sect. 4.1.

Our models are built using Word2Vec (Mikolov et al. 2013a), with its default settings: continuous bag-of-word (CBOW) implementation, frequency threshold of 5, 5-sized context window, negative sampling, 100 dimensions. These settings have proven to be efficient for most natural language processing tasks. Some studies opt for a higher frequency threshold, but it has not been shown yet that it significantly improves the results. In the present study, we choose to favor lexical diversity and large numbers of analyzed items.

Each tagged lemma with a frequency ≥ 5 in the corpus is represented by a vector in a model. A centroid of vectors can be computed by averaging these vectors using the formula (1). The neighborhood of a centroid can be outlined by computing its proximity score to every word in the model.

$$\vec{C} = \frac{1}{n} \sum_{i=1}^n \vec{v}_i \quad (1)$$

It has been shown lately that distributional semantics models suffer from intrinsic variability (Antoniak and Mimno 2018; Pierrejean 2020). The training of models by Word2Vec involves stochastic methods at various stages, which results in variation in vector representation from one model to another, even when these models are trained with the same parameters. Pierrejean and Tanguy (2018) state that those stochastic methods cause an average 17% variation among the 25 nearest neighbors of a given vector. To minimize this variability, we base our study on results from 5 distinct models built with the same settings. To the best of our knowledge, there is no recommended methodology with regard to the averaging of neighbors as part of a qualitative analysis. We choose to base our analysis on the nearest neighbors shared by the centroids in the 5 models. We compute in each model the centroid of prototypical ANs' vectors, and then compute the distance between that centroid and all other word vectors in the model. Proximity scores are averaged over the 5 models, and the nearest neighbors are selected for analysis, regardless of their POS tag.

4 Morphological diversity of agent nouns

In this section, we explore the morphological properties of ANs, by examining the vector space neighbors of French deverbal ANs ending in *-eur*. We first present the criteria used to select prototypical monosemous *-eur* ANs (Sect. 4.1). Then we observe that the nearest neighbors to the *-eur* ANs' centroids are morphologically diverse (Sect. 4.2), and we gather evidence to support that these neighbors can actually be considered ANs (Sect. 4.3). The distributional profile of morphologically heterogeneous AN candidates is finally examined, and compared to that of different agentive and non-agentive human-denoting nouns (Sect. 4.4).

4.1 Selecting prototypical ANs in French

To analyze the distributional profile of ANs and build a vector representation of prototypical ANs, we select a set of monosemous deverbal ANs suffixed with *-eur/-euse/-rice* (henceforth *-eur*). On the one hand, *-eur* nouns are consensually regarded as the main ANs in French (Benveniste 1948; Anscombe 2001; Fradin and Kerleroux 2003; Sleeman and Verheugd 2004; Roy and Soare 2014; Huyghe and Tribout 2015, among others), and *-eur* is probably the most productive affix for creating ANs in French—especially if one considers low-frequency items and neologisms. On the other hand, monosemy is a crucial condition to avoid multiple meaning effects and ambiguous distributional representations that could be induced by polysemy. Indeed the semantic versatility of *-eur* is well known. Deverbal nouns ending in *-eur* can nominalize a variety of semantic roles:

- (7) a. AGENT: *entraîneur* ‘coach’
 b. INSTRUMENT: *aspirateur* ‘vacuum cleaner’
 c. EXPERIENCER: *admirateur* ‘admirer’
 d. POSSESSOR: *détenteur* ‘holder’
 e. RECIPIENT: *receveur* ‘recipient’
 f. STIMULUS: *inspireur* ‘inspiration’⁵

There are different ways of accounting for the versatility of *-eur*. One is to assume that *-eur* is fundamentally a polysemous device that allows for a multiplicity of pre-determined meanings, possibly by extension from one core meaning (see Booij’s 1986 analysis of Dutch *-er*). Another option is to argue that affix semantics is basically underspecified and that underspecification is resolved in context (Lieber 2016; Lieber and Andreou 2018). In the latter approach, the interpretation of a given *-eur* noun would be determined altogether by the semantics of the base verb, by context and by encyclopedic knowledge, but the suffix itself is regarded as semantically vague. Whatever the theoretical stance, it is important here to distinguish between derivational semantics (i.e. semantic operations associated with word-formation processes) and lexical semantics (i.e. the semantics of words as specified in the lexicon). In any case, semantic representations have to account for the actual meaning of lexical items. The derivational semantics of *-er/-eur* may be underspecified, especially as regards the agent/instrument distinction. As already suggested by Benveniste (1948:61), the agent or instrument interpretation of an *-eur* noun derived from an action verb is presumably contingent, and unpredictable for newly formed nouns. Still in the process of

⁵That versatility is, to some extent, comparable to that of English *-er*, for which authors have proposed a syntactic explanation known as the ‘external argument generalization’, which stipulates that derived nominals ending in *-er* denote the external argument of the base verbs (Keyser and Roeper 1984; Rappaport Hovav and Levin 1992; Cohen 2016, among others). However, the external argument generalization cannot fully account for *-er* derivation, since *-er* deverbal nouns can denote themes (e.g. *scratcher* ‘lottery ticket that is scratched’) or locations (e.g. *diner* ‘place to dine in’), as shown by Ryder (1999), Booij and Lieber (2004), Alexiadou and Schäfer (2010), among others. These interpretations are not attested for French *-eur*, but the external argument generalization might be challenged by other data, such as the fact that some instrument-denoting nouns ending in *-eur* can hardly be used as the subject of the base verb (e.g. **Des écouteurs écoutent de la musique* ‘Headphones listen to music’, **Une liseuse lit des livres* ‘An e-reader reads books’).

lexicalization, meanings stabilize, and one can distinguish in French between lexicalized *-eur* nouns that strictly denote agents (8a), those that strictly denote instruments (8b), and those that denote both (8c).

- (8) a. AGENT NOUNS: *coiffeur* ‘hairdresser’, *acheteur* ‘buyer’, *déménageur* ‘mover’
 b. INSTRUMENT NOUNS: *réfrigérateur* ‘refrigerator’, *friteuse* ‘chip fryer’, *climatiseur* ‘air-conditioner’
 c. AGENT/INSTRUMENT NOUNS: *navigateur* ‘sailor’/‘navigator’, *semouse* ‘sower’/‘sowing machine’, *simulateur* ‘pretender’/‘simulator’

Nouns like *coiffeur*, *acheteur*, *déménageur* (8a) will not be used to refer to instruments used in hairdressing, buying or moving processes, nor will nouns like *réfrigérateur*, *friteuse*, *climatiseur* (8b) be used to denote persons who refrigerate, fry or air-condition something.⁶ In other words, the range of interpretations of a given lexicalized *-eur* noun is not context-dependent.

We assume in this study that semantic roles lexicalize in nominalizations and that French deverbal nouns suffixed with *-eur* have a fixed meaning as ‘agent’ and/or ‘instrument’ and/or other possible roles. Basing on that assumption, we build a list of monosemous ANs that will be analyzed in distributional semantics models. We first extract *-eur* nouns from the *lexique.org* database (New et al. 2004), which provides informations for more than 30,000 nouns in contemporary French. 2215 *-eur* nouns are listed, among which monosemous⁷ deverbal ANs are manually selected according to the following conditions. We exclude from the list:

- underived nouns (*peur* ‘fear’);
- feminine nouns ending in *-eur* that denote properties (*blancheur* ‘whiteness’) or phenomena (*clameur* ‘clamour’);
- non-deverbal nouns suffixed with *-eur* (*ambassadeur* ‘ambassador’);
- deverbal nouns with at least one non-agentive meaning, i.e. monosemous or polysemous nouns with one instrument, experiencer, possessor, recipient, stimulus, etc. meaning (*ventilateur* ‘fan’, *possesseur* ‘owner’, *sondeur* ‘pollster’/‘depth finder’, *tailleur* ‘tailor’/‘suit’);
- nouns with at least one meaning that is strongly desemanticized with respect to the base verb (*chauffeur* ‘driver’/‘chauffer’ ‘heat up’);
- nouns ending in *-euse* that are ambiguous between the feminine of an *-eur* noun and the feminine of a qualifying *-eux* noun (*chatouilleuse* ‘tickling woman’/‘ticklish woman’).

⁶It might be easier to exceptionally use instrument nouns in an agent interpretation than the opposite (as pointed out by Booij 1986 for Dutch *-er* nouns). Still in many cases such a use appears to be highly unlikely. As an indication, we examined in the French Wikipedia corpus the 1203 occurrences of *réfrigérateur* ‘refrigerator’, 82 occurrences of *friteuse* ‘chip fryer’, 195 occurrences of *climatiseur* ‘air-conditioner’, 40 occurrences of *ponceuse* ‘sander’, 19 occurrences of *bétonneuse* ‘concrete mixer’, 46 occurrences of *compacteur* ‘compact’, and observed that none was used to denote an agent.

⁷We define as monosemous ANs either nouns with only one lexical meaning, that meaning being agentive (e.g. *acheteur* ‘buyer’), or nouns with several meanings, all of which are agentive (e.g. *ouvreur* ‘opener’/‘usher’). Monosemy is assessed by using three lexicographic resources: *Le Trésor de la Langue Française Informatisé*, *Le Petit Robert de la langue française* and *Wiktionnaire*.

We do not a priori dismiss the possibility that non-deverbal nouns can be ANs, but since our purpose is precisely to investigate the agentivity of nouns that are not prototypical ANs, we do not include such candidates in our list. Nouns morphologically related but semantically unrelated to verbs (e.g. *chauffeur* ‘driver’) are excluded because their morphosemantic unanalyzability cannot guarantee agentivity. Since the distinction between agents and instruments is debatable in some cases, we exclude by default any noun that can be analyzed as instrument-denoting (i.e. that fundamentally corresponds to Z rather than X in *X Verb Y with Z*, see Sect. 2.1). On the opposite, we include nouns that denote unintentional agents (*ronfleur* ‘snorer’) or non necessarily animate agents (*perturbateur* ‘disrupter’), as well as nouns that ambiguously derive from verbal or nominal bases (*boxer* ‘box’/ *boxe* ‘boxing’ → *boxeur* ‘boxer’). Our decision in the latter case is based on the prototypicality of verbal derivation for *-eur*, and on the fact that the relation between possibly ambiguous *-eur* nouns and their potential base verbs is both transparent and similar to the one that applies to unambiguously deverbal *-eur* ANs.

1121 ANs are kept from the original list of 2215 *-eur* nouns. 681 out of those 1121 ANs occur in the French Wikipedia with a frequency ≥ 5 . It appears that the selected nouns almost exclusively denote humans, with the exception of a few nouns that denote animals (*rongeur* ‘rodent’) or chemical agents (*inhibiteur* ‘inhibitor’), and a few underspecified nouns with regard to animacy (*catalyseur* ‘catalyst’). These 681 ANs constitute the initial target words (that we will refer to as ‘seeds’) used to analyze the distribution of ANs in the corpus. We only take into account seeds labeled as common nouns by Talismane tagger, so as to exclude POS ambiguity (e.g. between noun and adjective, as in *enchanteur* ‘enchanter’/‘enchanting’). A centroid is computed by averaging the vectors of the 681 nouns, in 5 models trained with the same parameters. The nearest neighbors to the 5 centroids are identified by averaging their proximity scores to these centroids. For analysis purposes, the 100 nearest neighbors are scrutinized. Their examination is presented in the next section.

4.2 Lexical neighborhood of *-eur* ANs’ centroids

We carry out a morphosemantic analysis of the 100 nearest neighbors to the *-eur* ANs’ centroids (see Table 1 for a sample of them).⁸ First, it can be noted that these neighbors are all nouns, and that they denote human beings, the only exception being *chien* ‘dog’ in the 61st rank. Second, 27 of the 100 nearest neighbors correspond to seeds, i.e. they are part of the 681 original nouns that were used to compute the centroids. These nouns do not necessarily appear in the first ranks: the first ones are *coiffeur* ‘hairdresser’ in the 6th rank, *soigneur* ‘healer’ in the 14th rank, *colporteur* ‘peddler’ in the 19th rank, and the last ones are *ferrailleur* ‘scrap dealer’ in the 99th rank, *masseur* ‘masseur’ in the 94th rank, and *tricheur* ‘cheater’ in the 89th rank. Overall the 27 seeds distribute regularly among the 100 nearest neighbors. They do not correspond to the very core neighborhood of the centroid, as one might have expected, but are merely found among other vectors. Third, the morphological properties of the neighbors are quite diverse. The 100 nearest neighbors include:

⁸Detailed results can be viewed online at <https://github.com/french-agent-nouns/data>.

Table 1 20 nearest neighbors to the *-eur* ANs' centroids and average proximity score

Neighbor		Prox.	Neighbor		Prox.
<i>plombier</i>	'plumber'	0.769	<i>camionneur</i>	'truck driver'	0.699
<i>truand</i>	'crook'	0.751	<i>pickpocket</i>	'pickpocket'	0.698
<i>escroc</i>	'swindler'	0.745	<i>cuisinier</i>	'cook'	0.697
<i>rabatteur</i>	'beater'/'reaping reel'	0.730	<i>soigneur</i>	'healer'	0.693
<i>proxénète</i>	'pimp'	0.722	<i>charlatan</i>	'charlatan'	0.692
<i>coiffeur</i>	'hairstylist'	0.718	<i>magicien</i>	'magician'	0.691
<i>prestidigitateur</i>	'prestidigitator'	0.717	<i>voleur</i>	'thief'	0.690
<i>garagiste</i>	'mechanic'	0.704	<i>voyou</i>	'lout'	0.687
<i>gangster</i>	'gangster'	0.704	<i>colporteur</i>	'hawker'	0.683
<i>malfrat</i>	'thug'	0.699	<i>cambricoleur</i>	'housebreaker'	0.682

Table 2 Morphological properties of the neighbors of the *-eur* ANs' centroids (/100)

Affixed	Converted	Compound	Simple	Opaque	Clipping	Indeterminate
64	8	3	10	8	1	6

- affixed nouns (*cuisinier* 'cook');
- converted nouns (*drogué* 'drug addict');
- compounds (*photographe* 'photographer');
- morphologically simple nouns (*proxénète* 'pimp');
- opaque nouns, i.e. nouns whose form is only partly analyzable (*ivrogne* 'drunkard');
- clippings (*indicateur* 'informer' → *indic*);
- indeterminate nouns (*assassin* 'murderer').

We annotate as 'indeterminate' nouns that do not have any morphological exponent but are morphologically and semantically related to an existing non-nominal lexeme (*assassin* 'murderer'/*assassiner* 'murder'), assuming with Tribout (2010) that the direction of conversion cannot be decided in such cases. 'Indeterminate' lexemes are usually indeterminate between converted and morphologically simple nouns (provided they are not derived from any third lexeme). Note that the morphological analysis we provide is based on synchrony and does not take into account historical word-formation. In case of multiple formation, the annotation only applies to the last morphological operation. For instance, *couturier* 'fashion designer' is analyzed as a denominal noun suffixed with *-ier* and based on *couture* 'fashion design', though *couture* itself is derived from *coudre* 'sew'.

The morphological type of the 100 neighbors (i.e. affixed, converted, compound, etc.) is presented in Table 2. 64 of these neighbors are affixed nouns,⁹ using a variety

⁹These 64 neighbors include 63 suffixed nouns and 1 prefixed noun (*contremaître* 'foreman'). The possible agentivity of the latter is presumably inherited from its base noun. It can hardly be speculated that *contre-* as a prefix forms ANs, contrary to what is the case for suffixes *-ier*, *-iste*, *-ard*, etc.

Table 3 Suffix of the derived neighbors of the *-eur* ANs' centroids (/100)

<i>-eur</i>	<i>-ier</i>	<i>-iste</i>	<i>-ard</i>	<i>-on</i>	<i>-aire</i>	<i>-ien</i>
33	16	7	3	2	1	1

Table 4 Base POS of the derived neighbors of the *-eur* ANs' centroid (/100)

Verb	Noun	Adjective
37	29	6

of suffixes including:

- *-eur* (*balayeur* ‘sweeper’);
- *-ier* (*bijoutier* ‘jeweler’);
- *-iste* (*marionnettiste* ‘puppeteer’);
- *-ard* (*motard* ‘motorcyclist’);
- *-on* (*forgeron* ‘blacksmith’);
- *-aire* (*faussaire* ‘counterfeiter’);
- *-ien* (*magicien* ‘magician’).

The distribution of these suffixes among the 100 first neighbors is shown in Table 3. It can be noted that 7 *-eur* nouns in the neighborhood list are not seeds, which is due to the fact that some of them are denominal (*camionneur* ‘truck driver’, *farceur* ‘prankster’, *cascadeur* ‘stuntman’), some of them do not have a free verbal base in French (*prestidigitateur* ‘prestidigitator’, *délateur* ‘informer’, *imposteur* ‘fraud’),¹⁰ and one is polysemous between an agentive and an instrumental meaning (*rabatteur* ‘beater’/‘reaping reel’)—although in that particular case, the agentive meaning is overwhelmingly found in the Wikipedia corpus.

As can be seen in Table 4, the bases of the derived (i.e. affixed and converted) neighbors are mostly verbs and nouns, with a few adjectives in case of conversion (*criminel* ‘criminal’, *sadique* ‘sadistic’, *alcoolique* ‘alcoholic’). In case the POS of the base is ambiguous, we align the analysis with the most represented word-formation patterns, as long as it is coherent with the semantic analysis of the morphological construction. For instance, we consider *fêtard* ‘partygoer’ to derive from the verb *fêter* ‘celebrate’ rather than from the noun *fête* ‘celebration’, given that there is very few, if any, *-ard* nouns derived from action nouns in French, as opposed to *-ard* deverbal nouns (*fuyard* ‘fugitive’, *vantard* ‘boaster’, *pleurnichard* ‘whiner’).

The bases of the derived neighbors are semantically heterogeneous, as shown in Table 5. They may denote actions (*course* ‘errand’ → *coursier* ‘courier’), objects (*machine* ‘machine’ → *machiniste* ‘stagehand’), properties (*sadique* ‘sadistic’ → *sadique* ‘sadist’), domains (*couture* ‘fashion design’ → *couturier* ‘fashion designer’)

¹⁰We take these nouns to be derived because their base can be identified in other analyzable words. *Prestidigitateur* ‘prestidigitator’, *délateur* ‘informer’, *imposteur* ‘fraud’ are morphosemantically congruent with the deverbal *-eur* ANs, and their possible base can be identified in *prestidigitation* ‘prestidigitation’, *délation* ‘informing’, *imposture* ‘sham’, which are themselves morphosemantically congruent with the deverbal *-ion* and *-ure* action nouns.

Table 5 Semantic base of the derived neighbors of the *-eur* ANs' centroids (/100)

Action	Object	Property	Domain	Institution
43	15	6	6	2

or institutions (*police* 'police' → *policier* 'policeman'). Adjectives are analyzed as property-denoting items. Verbs are analyzed as action-denoting or property-denoting items, depending on their dynamic/stative inherent aspect. Nouns are more diverse and can actualize the whole range of semantic types. Linguistic tests taken from the literature (Godard and Jayez 1993; Flaux and Van de Velde 2000; Huyghe 2015, among others) are used to support their semantic classification. Action-denoting nouns can be used as the subject of *avoir lieu* 'take place', *se produire* 'happen', or as the object of *effectuer* 'carry out', *accomplir* 'perform', *procéder à* 'proceed to'. Object-denoting nouns can be the subject of *se trouver* 'be' followed by a spatial locative. Property-denoting nouns can be used in *être d'un grand N* 'be of great N', *état de N* 'state of N', or as the object of *ressentir* 'feel', *éprouver* 'experience', *faire preuve de* 'show'. Domain-denoting nouns are compatible with the light verb construction *faire du N* 'do some N'. Institution-denoting nouns can be the subject of *être fondé* 'be founded' followed by a temporal locative, or can be used in expressions such as *être nommé à la tête du N* 'be appointed head of the N'. In case a base word is subject to polysemy or homonymy, we annotate the meaning that most closely matches the derived word, as long as it respects the semantics of the morphological process involved. For instance, we consider *farceur* 'joker' to derive from the action meaning of *farce* 'joke' rather than from its object meaning 'stuffing', basing on semantic correspondence and on the fact that denominal *-eur* nouns can stem from action nouns (*cascadeur* 'stuntman', *bienfaiteur* 'benefactor', *navetteur* 'commuter').

4.3 Semantic analysis

At this stage, we cannot infer from the analysis of *-eur* ANs' neighbors any information about nominal agentivity. In particular, we cannot conclude that the nearest neighbors to the *-eur* ANs' centroids are themselves ANs. Indeed, almost all *-eur* ANs used to compute these centroids combine essentially two semantic features: human and agent. The relative influence of these two features on the distribution of *-eur* ANs is unknown.

To assess whether *-eur* ANs' centroids are sensitive to agentivity, we calculate their proximity to non-agentive human-denoting nouns. Here we consider general and phasic nouns (*personne* 'person', *vieillard* 'old man'), relational nouns (*fil*s 'son', *otage* 'hostage'), and demonyms (*Fidjien* 'Fijian', *Genevois* 'Genevan'). General and phasic nouns denote human beings without any further specification (Halliday and Hasan 1976; Mahlberg 2005), or only with the indication of age (Aleksandrova 2013). Some of them can be seen as hypernyms of other human-denoting nouns. We group general and phasic nouns together because of the fuzzy boundary between the two classes—mostly due to the lack of criteria for delineating the class of general nouns.

Table 6 Proximity scores and ranking of the 10 nearest general/phasic human nouns to the *-eur* ANs' centroids

General/Phasic noun		Prox.	Rank
<i>homme</i>	'man'	0.616	118
<i>adolescent</i>	'teenager'	0.577	240
<i>gars</i>	'guy'	0.566	281
<i>gens</i>	'people'	0.551	359
<i>bambin</i>	'toddler'	0.550	365
<i>quadragénaire</i>	'quadragenarian'	0.531	521
<i>vieillard</i>	'old man'	0.520	604
<i>sexagénaire</i>	'sexagenarian'	0.519	622
<i>garçonnet</i>	'little boy'	0.493	932
<i>quinquagénaire</i>	'quinquagenarian'	0.473	1194

As for relational nouns, they denote human beings in an interpersonal relation with others, and are compatible with genitives denoting the related persons (Vikner and Jensen 2002; Partee and Borschev 2003; Barker 2008, among others). Demonyms denote inhabitants and are derived from place names. We compile lists of non-agentive human nouns from two existing resources: the *Humanymes* database¹¹ for general, phasic and relational nouns, and *Prolexbase*¹² for demonyms. The lists are manually filtered so as to remove polysemous nouns, and completed with a selection of synonyms taken from the *Dictionnaire Electronique des Synonymes* edited by the CRISCO lab.¹³ Eventually 46 general/phasic nouns, 84 relational nouns, and 18080 demonyms are kept in the lists. Among them, respectively 39, 65 and 380 nouns occur in the French Wikipedia with a frequency ≥ 5 .

Tables 6, 7 and 8 present for each class of human nouns the 10 nearest nouns to the *-eur* ANs' centroids over the 5 models. The proximity scores are much lower than in the case of the first 100 neighbors previously analyzed. General/phasic nouns, relational nouns and demonyms have a low average neighborhood ranking to the *-eur* ANs' centroids.¹⁴

The fact that non-agentive human nouns do not feature among the nearest neighbors to the *-eur* ANs' centroids, as opposed to clearly agentive nouns (in the case of the 27 seeds found among the 100 nearest neighbors), supports the hypothesis that *-eur* ANs' centroids are semantically determined by agentivity. We will assume that the proximity to these centroids reveals a semantic combination of both human and agentive features, and can thus be used as an indicator of agent meaning.

¹¹<https://humanymes.u-strasbg.fr/>.

¹²<https://www.cnrtl.fr/lexiques/prolex/>.

¹³<https://crisco2.unicaen.fr/des/>.

¹⁴The general noun *homme* 'man' is an exception, since it appears in the 118th rank, with an average proximity score of 0.616 to the *-eur* ANs' centroids. That proximity could be explained by the fact that *homme* is a hypernym of other human nouns—given that hyponymy is a factor for distributional similarity (Mikolov et al. 2013b)—and by the fact that it is part of several multiword expressions that denote agents (e.g. *homme de main* 'henchman', *homme d'entretien* 'maintenance man').

Table 7 Proximity scores and ranking of the 10 nearest relational human nouns to the *-eur* ANs' centroids

Relational noun		Prox.	Rank
<i>copain</i>	'friend/boyfriend'	0.591	191
<i>amant</i>	'lover'	0.579	231
<i>invité</i>	'guest'	0.549	376
<i>compagnon</i>	'partner'	0.541	427
<i>camarade</i>	'friend/classmate'	0.528	544
<i>colocataire</i>	'roommate'	0.526	563
<i>fiancé</i>	'fiancé'	0.512	686
<i>père</i>	'father'	0.492	936
<i>rejeton</i>	'offspring'	0.486	1016
<i>frère</i>	'brother'	0.485	1018

Table 8 Proximity scores and ranking of the 10 nearest demonyms to the *-eur* ANs' centroids

Demonym		Prox.	Rank
<i>Cauchois</i>	'inhabitant of the Pays de Caux'	0.434	1981
<i>Tourquennois</i>	'inhabitant of Tourcoing'	0.422	2314
<i>béké</i>	'white Antilles citizen'	0.410	2732
<i>Berlinois</i>	'Berliner'	0.397	3278
<i>Asiatique</i>	'Asian'	0.397	3300
<i>Niçois</i>	'inhabitant of Nice'	0.391	3568
<i>Ardenmais</i>	'inhabitant of the Ardennes'	0.369	4925
<i>Véronais</i>	'Veronese'	0.367	5091
<i>Lorientais</i>	'inhabitant of Lorient'	0.367	5093
<i>Toulousain</i>	'Toulousian'	0.363	5441

It follows that the neighbors analyzed in Sect. 4.2 can be regarded as a representative sample of ANs. We conclude that French ANs are morphologically diverse as regards (un)derivedness, affix selection, and base selection. The morphological heterogeneity of the very first neighbors to the *-eur* ANs' centroids (e.g. the 20 nearest neighbors listed in Table 1) goes in line with that conclusion.

A closer look at the list of neighbors reveals unexpected results. For instance, we find in that list nouns converted from past participle bases: *drogué* 'drug addict' in the 33rd rank, and *travesti* 'cross-dresser' in the 78th rank. Such converted nouns are expected to be patient nouns rather than ANs, since they usually correspond to the internal argument of transitive verbs (e.g. *condamné* 'convicted person', *blessé* 'injured person', *invité* 'guest'). Their presence in the close neighborhood of ANs nonetheless makes sense if one considers that *drogué* and *travesti* would rather be defined as the subject in a reflexive construction (9a)–(10a) than as the object in a transitive construction of the base verb (9b)–(10b).

- (9) a. *Un drogué est quelqu'un qui se drogue.*
 'A drug addict is somebody who takes drugs' (lit. 'who drugs himself')
- b. *?Un drogué est quelqu'un que l'on drogue.*
 'A drug addict is somebody whom one drugs'
- (10) a. *Un travesti est quelqu'un qui se travestit.*
 'A cross-dresser is somebody who cross-dresses'
- b. *?Un travesti est quelqu'un que l'on travestit.*
 'A cross-dresser is somebody whom one dresses as opposite sex'

Whether reflexivity markers are analyzed as arguments or as voice markers, subjects of reflexives are assumed to preserve the semantic role of the corresponding transitive subjects. In the case of *se droguer* 'take drugs' and *se travestir* 'cross-dress', subjects can be fundamentally regarded as agents. It appears here that non-canonically agentive forms can be used to create ANs under some specific conditions. This could illustrate the idea, suggested by Bauer et al. (2013), that some morphological constructions are primarily agentive whereas others are secondarily or even accidentally agentive—in which case the denotation of agents is not the main semantic function of the construction. The notion of “secondary usage” (Bauer et al. 2013:231) has to be clarified though, for it may refer to two distinct situations. Secondary agent meanings are mentioned for English *-ee* (*devotee*), *-ing* (*following*), *-ation* (*administration*), and for conversion (*cook*). It can be pointed out that in some cases, as opposed to others, derived nouns are ambiguous between an agent and a non-agent meaning. If polysemy is necessary, then the agent meaning could be created by metonymy, not by morphological construction. For instance, it seems that French collective ANs suffixed with *-ion* (*rédaction* 'editorial board', *rébellion* 'rebellion', *immigration* 'immigration') are always ambiguous between an agent and an action meaning. It could be argued that in such a case the agent meaning results from metonymy, and that *-ion* as an affix does not yield ANs, not even secondarily. Thus a distinction can be made between affixes associated with an agent-forming pattern—whether it is the most productive pattern associated with the affix or not—and affixes that happen to be found in ANs but are presumably not associated with any morphological agentive construction.

4.4 Non-*eur*-suffixed AN candidates

In order to confirm the morphological diversity of ANs, we examine the distributional profile of AN candidates that are not deverbal *-eur* nouns. We evaluate their position in vector space and their proximity to different agentive and non-agentive human nouns' centroids.

It should be stressed that the method we use here provides an indicator, but not a test to diagnose the agentivity of a given lexical item. By analyzing samples of items, we aim at general inferences concerning morphological types. Indeed the proximity of a given word to centroid A with comparison to centroid B is certainly a clue but not a condition for A vs. B semanticism. For instance, some *-eur* deverbal ANs (e.g. *guetteur* 'spotter', *randonneur* 'hiker', *sauveur* 'savior') are closer to non-agentive human nouns' centroids than to *-eur* ANs' centroids. Conversely, some non-agentive

human nouns (e.g. *Ardennais* ‘inhabitant of the Ardennes’, *gus* ‘guy’, *zig* ‘bloke’) are closer to *-eur* ANs’ centroids than to non-agentive ones. However, the general tendency is clearly that most *-eur* ANs are closer to agentive centroids than to non-agentive ones, whereas most non-agentive human nouns are closer to non-agentive centroids than to agentive ones.

To compute centroids, we use the lists of monosemous non-agentive human nouns presented in Sect. 4.3. As extensive as they are, these lists significantly differ in size. The smallest list, that of general/phasic nouns, consists of 39 items with a frequency ≥ 5 in the French Wikipedia, while the original list of deverbal *-eur* ANs includes 681 items. The impact of such a difference on the comparison between centroids being unknown, we sample the lists so as to have the same number of seeds for each centroid. Given that the frequencies of human nouns also differ considerably, we divide each list into 39 balanced groups with regard to frequency, and we randomly select one item in each group so that frequencies are similarly distributed within seed lists.

We then select 20 AN candidates for each morphological type that is to be tested. These types include:

- denominal nouns suffixed with *-eur* (*autostoppeur* ‘hitchhiker’, *basketteur* ‘basketball player’, *précepteur* ‘private tutor’);
- nouns suffixed with *-aire* (*bibliothécaire* ‘librarian’, *gestionnaire* ‘manager’, *plagiaire* ‘plagiarist’);
- nouns suffixed with *-iste* (*pianiste* ‘pianist’, *éclairagiste* ‘lighting engineer’, *exorciste* ‘exorcist’);
- nouns suffixed with *-ien* (*chirurgien* ‘surgeon’, *marathonien* ‘marathon runner’, *historien* ‘historian’);
- nouns suffixed with *-ier* (*braconnier* ‘poacher’, *caissier* ‘cashier’, *luthier* ‘stringed instrument maker’);
- nouns ending in *-ant*, possibly converted from present participles of verbs (*combattant* ‘fighter’, *manifestant* ‘demonstrator’, *surveillant* ‘monitor’);
- nouns in a conversion relation to verbs, but indeterminate as regards the direction of the conversion (*arbitre* ‘referee’, *pèlerin* ‘pilgrim’, *pilote* ‘pilot’);
- morphologically simple nouns (*médecin* ‘doctor’, *architecte* ‘architect’, *scribe* ‘scribe’).

The candidates are monosemous human nouns selected from different sources: the *Lexeur* database¹⁵ for non-deverbal *-eur* nouns; *lexique.org* for *-aire*, *-ant*, *-iste*, *-ien* and *-ier* nouns; Tribout (2010) for nouns in a conversion relation to verbs; Tribout et al. (2014) for morphologically simple nouns. The selected nouns are not part of the 100 nearest neighbors to the deverbal *-eur* ANs’ centroids previously analyzed. We limit the selection to 20 candidates of each morphological type because we do not aim at exhaustiveness, but favor a comparative overview that requires a fixed number of candidates of each type. Not all morphological types being equally represented in the corpus, we base our sample on the least represented type—in the present case simple nouns, for which monosemous AN candidates with an adequate frequency were not

¹⁵ Soon to be available at <http://redac.univ-tlse2.fr/lexiques/>.

Table 9 Sample of AN candidates and their proximity to *-eur* ANs', general/phasic human nouns', relational human nouns' and demonyms' centroids. The highest proximity score for each AN candidate is indicated in bold

AN candidate		<i>-eur</i> ANs	Gen./Phasic	Relational	Demon.
<i>basketteur</i>	'basketball player'	0.336	0.151	0.198	0.175
<i>précepteur</i>	'private tutor'	0.412	0.444	0.676	0.044
<i>bibliothécaire</i>	'librarian'	0.403	0.260	0.341	0.168
<i>gestionnaire</i>	'administrator'	0.366	0.121	0.072	0.098
<i>pianiste</i>	'pianist'	0.332	0.244	0.408	0.048
<i>exorciste</i>	'exorcist'	0.490	0.463	0.376	0.055
<i>chirurgien</i>	'surgeon'	0.559	0.418	0.408	0.065
<i>historien</i>	'historian'	0.477	0.265	0.294	0.027
<i>braconnier</i>	'poacher'	0.542	0.566	0.355	0.128
<i>horloger</i>	'clockmaker'	0.518	0.369	0.348	0.069
<i>manifestant</i>	'demonstrator'	0.332	0.385	0.087	0.059
<i>combattant</i>	'fighter'	0.600	0.428	0.256	0.330
<i>arbitre</i>	'referee'	0.483	0.274	0.152	0.367
<i>pilote</i>	'pilot'	0.601	0.229	0.246	0.395
<i>médecin</i>	'doctor'	0.561	0.505	0.485	0.030
<i>architecte</i>	'architect'	0.387	0.119	0.245	0.005

Table 10 Distribution of AN candidates according to their proximity to *-eur* ANs', general/phasic human nouns', relational human nouns' and demonyms' centroids. The most represented centroid proximity for each morphological type is indicated in bold

Morphological type	<i>-eur</i> ANs	Gen./Phasic	Relational	Demon.
Denominal <i>-eur</i>	13 (65%)	5 (25%)	–	2 (10%)
Suffixed with <i>-aire</i>	19 (95%)	–	1 (5%)	–
Suffixed with <i>-iste</i>	16 (80%)	3 (15%)	1 (5%)	–
Suffixed with <i>-ien</i>	16 (80%)	4 (20%)	–	–
Suffixed with <i>-ier</i>	14 (70%)	5 (25%)	1 (5%)	–
Ending in <i>-ant</i>	13 (65%)	4 (20%)	2 (10%)	1 (5%)
Unmarked conversion	9 (45%)	8 (40%)	3 (15%)	–
Simple	10 (50%)	9 (45%)	1 (5%)	–

numerous. Table 9 presents a sample of the average proximity scores calculated for AN candidates over 5 models. Table 10 presents the results for each morphological type.

It appears that many if not most candidates of each morphological type are closer to *-eur* ANs' centroids than to other human nouns' centroids. This result reinforces the claim that a large variety of morphological constructions can yield ANs, and that generally speaking, ANs in French are not limited to deverbal nouns ending in *-eur*.

5 Subclassification of agent nouns

As seen in Sect. 2, the homogeneity of the AN class can be investigated not only from a morphological, but also from a semantic point of view. The morphological diversity of ANs further calls for semantic investigation, and it can be asked whether it correlates with different types of agent meaning. In this section, we focus on AN subclassification and its morphological counterpart. In Sect. 5.1, we describe a sub-sampling of prototypical functional, occasional and behavioral *-eur* ANs. Section 5.2 presents some distributional evidence to support the distinction between the three subtypes. In Sect. 5.3, we show that this distinction extends to non-*-eur*-suffixed ANs, and that it most likely correlates with specific properties as regards morphological type, base selection and affix selection.

5.1 Subsamples of *-eur* ANs

Following the line adopted in the agent vs. instrument analysis (Sect. 4.1), we assume that ANs can be lexically defined with respect to the distinction between functional, occasional and behavioral ANs. To distinguish between the three types of ANs, we rely on the tests introduced in Sect. 2.3, and presented below as Conditions 1–3. These conditions are sufficient for subcategorizing ANs respectively as functional, occasional and behavioral ANs.

- Condition 1: the noun can be used as a bare predicate without any complement (*X est N* ‘X is a N’ lit. ‘X is N’);
- Condition 2: the noun is compatible with a specific complement that denotes a participant in a particular event (*le N de x* ‘the N of x/x’s N’);
- Condition 3: the noun is compatible with a size adjective in a non-intersective interpretation, without any other complement (*un gros N* ‘a big N’).

Of course, polysemy may occur. Nouns like *animateur* ‘facilitator’/‘presenter’, *racketteur* ‘racketeer’ and *inventeur* ‘inventor’ are, respectively, a functional and an occasional AN (11a), an occasional and a behavioral AN (11c), a functional, an occasional and a behavioral AN (11d).

- (11) a. *Pierre est {animateur / l’animateur de ce débat / #un gros animateur}*.
 ‘Pierre is {a facilitator/the moderator of this debate/a big presenter}’
 b. *Pierre est {??racketteur / le racketteur de Marie / un gros racketteur}*.
 ‘Pierre is {a racketeer/Mary’s racketeer/a big racketeer}’
 c. *Pierre est {inventeur / l’inventeur de ce dispositif / un gros inventeur}*.
 ‘Pierre is {an inventor/the inventor of this device/a big inventor}’

Coercion is also possible for occasional ANs that derive from transitive verbs. These can be interpreted as behavioral ANs when they are used with quantitatively unbounded complements, such as the indefinite plural complements in (12). The absence of a complement is required in Conditions 1 and 3 precisely to avoid such coercion effects.

- (12) *Cet homme est {un grand agresseur de personnes âgées / un gros envoyeur d’e-mails / un gros acheteur de voitures anciennes}*.
 ‘This man is {a great attacker of elderly people/a big email sender/a big buyer of vintage cars}’

Table 11 Identification of monosemous subtypes of deverbal ANs

	Cond. 1	Cond. 2	Cond. 3	Cond. 4
Monosemous functional ANs	1	0	0	1
Monosemous behavioral ANs	0	0	1	1
Monosemous occasional ANs	0	1	0	1
	0	0/1	0	0

To examine the distributional profile of functional, occasional and behavioral ANs, we randomly select 50 monosemous *-eur* ANs of each type, among the deverbal *-eur* ANs that are present in our models. Their identification is based on the three previously mentioned conditions, completed with a fourth one (Condition 4: the noun is derived from a transitive verb). That additional condition is needed to identify monosemous functional and behavioral ANs, because there is no single positive test to detect occasional ANs derived from intransitive verbs. It compensates for the fact that ANs derived from intransitive verbs and satisfying Condition 1 or 3 but not Condition 2 may be polysemous between functional or behavioral and occasional. In contrast, ANs that derive from intransitive verbs and do not satisfy Conditions 1 nor 3 are by default monosemous occasional ANs. Table 11 presents the diagnoses for functional, occasional and behavioral ANs that can be inferred from the four conditions.

We now address two questions about AN subclassification: does the distinction between functional, occasional and behavioral ANs (i) correlate with distributional differences between ANs, and (ii) extend to non-*eur* ANs, possibly in correlation with morphological properties of ANs? These questions are answered in the two following subsections.

5.2 Distributional relevance

To evaluate the distributional relevance of the distinction between functional, occasional and behavioral ANs, we first apply a clustering algorithm to the 150 monosemous ANs we sampled. In each of the 5 models used in our study, we operate a hard spherical k-means partition of the 150 ANs into 3 clusters, and evaluate the correspondence between these clusters and the three predefined classes. In Models 1 and 5, the majority of each subclass of ANs falls into different clusters. In Models 2, 3, and 4 respectively, most occasional and behavioral ANs, occasional and functional ANs, functional and behavioral ANs are grouped into the same cluster. In all 5 models, the Pearson's chi-squared test shows a significant relationship between the clustering and the distinction between functional, occasional and behavioral ANs (at $p < .05$).¹⁶

In order to refine these results, we conduct a cluster analysis of the subclasses two by two (i.e. hard spherical k-means partition of 100 objects into 2 classes). The application of the Pearson's chi-squared test with Yates' continuity correction to the

¹⁶Detailed results can be viewed online at <https://github.com/french-agent-nouns/data>.

Table 12 Overlap between functional, occasional and behavioral seeds and neighbors to the corresponding centroids (/100)

	Functional neighbors	Occasional neighbors	Behavioral neighbors
Functional seeds	11	2	0
Occasional seeds	0	7	0
Behavioral seeds	1	1	3

resulting data shows that the distinction between functional and behavioral ANs is significant in all 5 models, whereas the distinction between functional and occasional ANs is significant in 4 models, and the distinction between occasional and behavioral ANs is significant in 3 models (at $p < .05$).

Overall, the clustering results support the distinction between functional, occasional and behavioral ANs. They show differences in distinctiveness between the three subclasses, functional ANs being more clearly separated from the two other subclasses than are behavioral ANs, and then occasional ANs. Nevertheless, these specificities do not invalidate the distributional relevance of the three-way distinction, which can be assumed to be (at least to some extent) determined by the semantic distinction between the three subclasses. To support and build on that assumption, we compute functional, occasional and behavioral ANs' centroids from our subsamples of ANs, and scrutinize their neighborhood in 5 models. The 100 nearest neighbors to each type of AN's centroids are identified by averaging their proximity scores to the centroids over the 5 models. In line with the idea that these centroids are representative of AN subtypes, it can be noted that the overlap rate between seeds and nearest neighbors of different subtypes is very low if not zero, contrary to that between seeds and nearest neighbors of the same subtype, as shown in Table 12.

The relevant neighbors among the nearest neighbors to functional, occasional and behavioral ANs' centroids are also analyzed, with respect to Conditions 1–3 previously used to distinguish between the three categories. We define 'relevant neighbors' as the neighbors that are both nouns and clearly not non-candidate for agentivity, i.e. we exclude words that are not automatically tagged as nouns (e.g. adjectives *vani-teux* 'conceited', *cupide* 'greedy', *fourbe* 'deceitful' in the neighborhood of behavioral ANs' centroids), as well as nouns that cannot denote agents—in the present case eventuality-denoting nouns (e.g. *ingratitude* 'ungratefulness' in the neighborhood of occasional ANs' centroids) and strictly relational nouns (e.g. *proche* 'close family/friend' in the neighborhood of occasional ANs' centroids). The number of relevant neighbors among the 100 nearest ones is: 100 in the case of functional ANs, 90 in the case of occasional ANs, and 79 in the case of behavioral ANs. As stated before, Conditions 1–3 are used to identify respectively functional, occasional and behavioral ANs, but polysemy between different agentive meanings frequently occurs. The results of the analysis are given in Table 13.

The distribution between the relevant neighbors that satisfy Conditions 1–3 is statistically dependent on the centroids' neighborhood to which they belong ($\chi^2(4, N = 295) = 168.7973, p < 2.2e-16$). Significantly more neighbors of one subclass than

Table 13 Relevant neighbors of functional, occasional and behavioral ANs' centroids satisfying Conditions 1–3 (/100)

	Cond. 1	Cond. 2	Cond. 3
Functional neighbors	98	6	9
Occasional neighbors	24	29	43
Behavioral neighbors	12	3	71

of the other two satisfy a condition for the classification in that particular subclass.¹⁷ These results are congruent with the idea that the distributional differences between functional, occasional and behavioral ANs correlate with the difference between the three agentive subclasses. In other words, the semantic distinction between these subclasses determines (at least to some extent) the distributional profile of ANs.

Two observations corroborate that claim. First, the neighborhood of the behavioral ANs' centroids contains 18 adjectives (vs. zero in the case of functional and occasional ANs' centroids), as well as many nouns that are converted from or frequently used as adjectives (e.g. *maniaque* 'maniac'/'obsessive', *sadique* 'sadist'/'sadistic', *teigneux* 'nasty person'/'nasty', *débrouillard* 'resourceful person'/'resourceful', *rêveur* 'dreamer'/'dreamy', *ronchon* 'grump'/'grumpy').¹⁸ This can be explained by the fact that behavioral ANs describe agentive attitudes that are associated with habits and general ways of behaving. They are close to property-denoting words, being used to characterize referents in a way that is comparable to that of adjectives. Second, the 10 non-relevant neighbors in the neighborhood of the occasional ANs' centroids are nouns, among which 5 have at least one eventive meaning (vs. zero in the case of functional and behavioral ANs' centroids), i.e. they are compatible with *avoir lieu* 'take place' or *se produire* 'happen' and can denote the occurrence of an event. These nouns (*agissement* 'act', *infortune* 'misfortune', *malheur* 'misfortune', *manigance* 'scheme', *méfait* 'misdeed'), whether they involve an agent or not, share with occasional ANs the reference to a particular event, which could be an explanation for their presence in the neighborhood of occasional ANs' centroids—if that neighborhood, as we argue, relates to the denotation of particular occurrences of actions.

Some additional remarks can be made concerning the subclass of functional ANs. As a confirmation of its distributional homogeneity, the neighborhood of functional ANs' centroids appears to be both more cohesive and more distant from others than

¹⁷Neighbors of occasional ANs' centroids are more regularly distributed than neighbors of the functional and behavioral ANs' centroids with respect to Conditions 1–3, which is consistent with our earlier finding that occasional ANs are less homogeneous in linguistic distribution than functional and behavioral ANs. 29 neighbors satisfying Condition 2 may seem rather few in the case of occasional neighbors. It has to be reminded though that Condition 2, as opposed to Conditions 1 and 3 with respect to functional and behavioral ANs, is not necessary for the classification as occasional AN (e.g. occasional ANs derived from intransitive verbs are likely not to satisfy that condition, as pointed out in Sect. 5.1). Still, the fact remains that significantly more neighbors of occasional ANs' centroids than of functional and behavioral ANs' centroids satisfy Condition 2, which suggests an overall greater correspondence to occasional meaning.

¹⁸The formal identity between nouns and adjectives might create issues in POS tagging. We did not investigate that point, and strictly relied on the POS tags provided by the tagger we used, which was coherent with our initial decision to analyze vectors of words automatically tagged as nouns.

Table 14 Overlap between the neighbors of functional, occasional and behavioral ANs' centroids (/100)

	Funct. neighbors	Occ. neighbors	Behav. neighbors
Functional neighbors	–	3	2
Occasional neighbors	3	–	16
Behavioral neighbors	2	16	–

are the neighborhoods of occasional and behavioral ANs' centroids. Non-relevant neighbors are not found in the neighborhood of functional ANs' centroids, contrary to what is observed in the case of the two other subclasses. Furthermore, the average proximity scores over 5 models between functional and occasional ANs' centroids (0.637) and functional and behavioral ANs' centroids (0.629) are lower than between occasional and behavioral ANs' centroids (0.702). That distance is confirmed by the overlap differences between the nearest neighbors of each subclass presented in Table 14. Observations converge to show that functional ANs form the most distinctive subclass of ANs.

5.3 Non-*eur*-suffixed functional, occasional and behavioral ANs

Another question we address is whether the distinction between the functional, occasional and behavioral AN subclasses extends to non-*eur*-suffixed ANs. A first clue here is that the relevant neighbors of the functional, occasional and behavioral ANs' centroids have various morphological profiles, many of which were already observed among the 100 nearest neighbors to the general *-eur* ANs' centroids. For instance, denominal or deverbal nouns suffixed with *-iste* (*trapéziste* 'trapeze artist'), *-ier* (*braconnier* 'poacher'), *-aire* (*manutentionnaire* 'warehouseman'), *-ard* (*débrouillard* 'resourceful person') or *-ien* (*comédien* 'actor') are present in the near neighborhood of functional, occasional and behavioral ANs' centroids. Moreover, some of the neighbors with these morphological profiles overlap with the neighbors of the general *-eur* ANs' centroids, thus qualifying for agentivity. Examples of such overlapping neighbors are given in (13).

- (13) a. AFFIXED NOUNS: *plombier* 'plumber' (G1, F2),¹⁹ *arriviste* 'social climber' (G62, B15), *faussaire* 'counterfeiter' (G34, O43)
 b. CONVERTED NOUNS: *drogué* 'drug addict' (G32, F100), *criminel* 'criminal' (G31, O18)
 c. COMPOUNDS: *photographe* 'photographer' (G59, F30), *ventriloque* 'ventriloquist' (G35, F51)
 d. MORPHOLOGICALLY SIMPLE NOUNS: *voyou* 'lout' (G18, B35, O91), *bandit* 'bandit' (G46, O40), *héros* 'hero' (G88, O10)
 e. OPAQUE NOUNS: *ivrogne* 'drunkard' (G39, O58, B10), *apprenti* 'apprentice' (G73, F9)

¹⁹Numbers in parentheses indicate the rank a noun occupies in a given averaged neighborhood. Neighborhoods are identified with letters F, O, B, G, respectively for functional, occasional, behavioral and general *-eur* ANs' neighborhoods.

Assuming that functional, occasional and behavioral ANs' centroids are indicative of the semantic distinction between the three subclasses, and taking into consideration the fact that the neighbors of these centroids heterogeneously satisfy Conditions 1–3 (see Table 13), it can be argued that some of these neighbors are indeed functional, occasional and behavioral ANs, and therefore that the three classes extend beyond the mere *-eur* ANs. Instances of functional, occasional and behavioral non-*eur* candidates included in the corresponding neighborhoods are listed respectively in (14), (15) and (16).

- (14) FUNCTIONAL ANS' NEIGHBORS: *plombier* 'plumber' (F2), *garagiste* 'mechanic' (F3), *vétérinaire* 'veterinarian' (F13), *électricien* 'electrician' (F20), *couturier* 'fashion designer' (F31), *comptable* 'accountant' (F35), *laborantin* 'laboratory technician' (F50), *maréchal-ferrant* 'farrier' (F42), *chirurgien* 'surgeon' (F48), *assistant* 'assistant' (F59), *proxénète* 'pimp' (F66), *taxidermiste* 'taxidermist' (F71), *radiologue* 'radiologist' (F75), *surveillant* 'monitor' (F88)
- (15) OCCASIONAL ANS' NEIGHBORS: *assassin* 'murderer' (O1), *meurtrier* 'murderer' (O2), *tortionnaire* 'torturer' (O5), *traître* 'traitor' (O9), *criminel* 'criminal' (O18), *conspirateur* 'conspirator' (O21), *coupable* 'culprit' (O34), *fraudeur* 'fraudster' (O37), *commanditaire* 'sponsor' (O38), *voleur* 'rapist' (O39), *opresseur* 'oppressor' (O54), *gêneur* 'troublemaker' (O79), *fugitif* 'fugitive' (O96), *incendiaire* 'arsonist' (O100)
- (16) BEHAVIORAL ANS' NEIGHBORS: *filou* 'rascal' (B2), *fanfaron* 'braggart' (B5), *vantard* 'boaster' (B6), *arriviste* 'social climber' (B15), *goinfre* 'glutton' (B17), *fainéant* 'idler' (B18), *bavard* 'chatterbox' (B23), *débrouillard* 'resourceful person' (B33), *mythomane* 'mythomaniac' (B40), *poivrot* 'boozer' (B43), *débauché* 'debauchee' (B68), *ronchon* 'grump' (B81), *goujat* 'boor' (B96), *fayot* 'toady' (B98)

An additional remark concerns moral connotation in occasional and behavioral ANs' neighbors. Whereas functional ANs' neighbors are generally free from connotation, occasional ANs' neighbors include many nouns that refer to harmful actions (i.e. criminal actions or attacks on others), and behavioral ANs' neighbors include many nouns that describe tendencies to act in a faulty or socially depreciated way. The same property can be found in some occasional seeds (e.g. *agresseur* 'attacker', *kidnappeur* 'kidnapper', *saboteur* 'saboteur') and behavioral seeds (e.g. *baratineur* 'smooth talker', *emmerdeur* 'pain in the ass', *magouilleur* 'wheeler-dealer'), but it is more prominent among neighbors. If one considers the nouns in (15) and (16) to be respectively occasional and behavioral ANs, then a recurrent lexical association appears between the denotation of occasional and behavioral agents on one hand and the reference to harmful actions/moral faults on the other. Although denominations for occasional and behavioral "positive" or "neutral" agents exist in French (e.g. *bienfaiteur* 'benefactor', *mangeur* 'eater'), these could be a minority or a less salient group among occasional and behavioral ANs. It can be noted that moral connotation is also quite salient, even though to a lesser extent, in the near neighborhood of the general *-eur* ANs' centroids (see Table 1), which could be explained by the presence of mixed functional, occasional and behavioral ANs among the general *-eur* seeds—occasional and behavioral seeds being responsible for morally connoted neighbors, and functional seeds for more neutral ones.

Table 15 Morphological type of the relevant functional, occasional and behavioral ANs' neighbors (/100)

	Affixed	Conv.	Compound	Simple	Complex	Redup.	Indet.
Functional	76	6	8	3	6	0	1
Occasional	39	15	1	19	6	0	10
Behavioral	22	9	4	13	4	1	26

On the assumption that the distinction between functional, occasional and behavioral agents applies to non-*eur* ANs, we wonder if it correlates with morphological properties of ANs. The morphosemantic analysis of functional, occasional and behavioral ANs' neighbors highlights some significant tendencies in that respect, and allows for several hypotheses about the correlation between the morphological properties of ANs and their semantic subclassification.

As far as morphological construction is concerned, the analysis of the relevant neighbors of functional, occasional and behavioral ANs' centroids shows that (i) affixation is more prominent in the neighborhood of functional ANs than in that of occasional and behavioral ANs, (ii) morphologically simple nouns are preferentially occasional and behavioral ANs' neighbors, and (iii) morphologically indeterminate constructions dominate in the neighborhood of behavioral ANs (see Table 15). The last point deserves a comment. Following the line of analysis laid out in Sect. 4.2, we annotate as 'indeterminate' nouns that are in a conversion relation with another word while the direction of conversion cannot be decided on a morphological basis. Out of the 26 indeterminate behavioral ANs' neighbors, 24 are in a conversion relation to an adjective (e.g. *pervers* 'pervert'/'perverse'), and 2 to a verb (e.g. *escroc* 'swindler'/'lescroquer' 'swindle'). Out of the 10 indeterminate occasional ANs' neighbors, 7 are in a conversion relation to an adjective (e.g. *scélérat* 'villain'/'villainous'), and 3 to a verb (e.g. *assassin* 'murderer'/'assassiner' 'murder'). When taking into account the fact that the 9 converted behavioral ANs' neighbors are all converted from adjectives, it appears that the most represented morphological profile among the relevant behavioral ANs' neighbors is noun in a conversion relation to an adjective, which is a distinctive feature of that neighborhood. Conversely, since out of the 15 converted occasional ANs' neighbors, 10 are converted from adjectives and 5 from verbs, most neighbors in a conversion relation to an adjective fall into the behavioral subclass.

There is indeed a significant dependence between the different neighborhoods and the POS of the base, as can be seen in Table 16 ($\chi^2(4, N = 167) = 45.176$, $p < .00001$). Nominal bases massively derive functional ANs' neighbors, which can be explained by the fact that nominal bases are prone to denote objects or domains, and that these are privileged semantic bases for functional ANs (see below). Occasional ANs' neighbors preferentially select verbal bases, which may have to do with the fact that verbs are the most likely to denote events. As for behavioral ANs' neighbors, if one considers that in addition to the 31 items in Table 16, 24 are possibly derived from adjectives (vs. 7 in the case of occasional ANs' neighbors and zero in the case of functional ANs' neighbors), then a potential correlation appears between adjectival bases and the neighborhood of behavioral ANs.

Table 16 Base POS of the derived relevant functional, occasional and behavioral ANs' neighbors (/100)

	Noun	Verb	Adjective
Functional	52	29	1
Occasional	12	31	11
Behavioral	3	19	9

Table 17 Semantic base of the derived relevant functional, occasional and behavioral ANs' neighbors (/100)

	Action	Object	Property	Domain	Institution
Functional	31	36	2	12	1
Occasional	37	5	11	2	0
Behavioral	20	1	10	1	0

As concerns the semantic type of the base, occasional and behavioral ANs' neighbors are preferentially derived from words that denote actions or properties, whereas object- or domain-denoting bases rather derive functional ANs' neighbors, as shown in Table 17. In the case of occasional ANs' neighbors, action-denoting bases clearly prevail. The reference to particular events in occasional ANs may often be dependent on the existence of a morphological base that can denote occurrences of actions. The prevalence of actions over properties is less obvious in the case of behavioral ANs. Should morphologically indeterminate neighbors be considered deadjectival, behavioral ANs' neighbors would mostly be derived from property-denoting bases. That indeterminacy somehow supports the view that behavioral ANs are semantically hybrid between action and property. As for functional ANs' neighbors, they generally denote professions or specialists, some of which can be defined by the production or handling of objects, or by activities associated with knowledge or practice domains—hence the selection of object- and domain-denoting bases. In both cases, the semantic action component is not inherited, but directly engendered in the semantic structure of derived ANs, in relation to the referent of the base.

Some correlations between affix selection and functional, occasional and behavioral ANs' neighborhood can be identified as well. Table 18 presents the distribution of the suffixes used in the nearest derived functional, occasional and behavioral ANs' neighbors. It appears that nouns ending in *-iste*, *-ier*, *-ien*, *-logue* tend to be functional ANs' neighbors, whereas *-aire* rather forms occasional ANs' neighbors, and *-ard* behavioral ANs' neighbors. Considering that functional, occasional and behavioral ANs' neighbors are likely to be respectively functional, occasional and behavioral ANs, it can be hypothesized that agentive suffixes have predilections as to which subtype of ANs they form. Possible correlations have to be further studied though, and larger samples of ANs are needed to formulate reliable generalizations.

Table 18 Suffix of the derived relevant functional, occasional and behavioral ANs' neighbors (/100)

	<i>-eur</i>	<i>-ier</i>	<i>-iste</i>	<i>-ien</i>	<i>-aire</i>	<i>-ard</i>	<i>-logue</i>	<i>-on</i>
Functional	24	31	11	6	1	0	2	1
Occasional	26	5	0	1	4	0	0	0
Behavioral	15	0	1	1	0	5	0	0

6 Conclusion

In this article, we have argued that nominal agentivity could be captured by distributional semantics analysis. The proximity to prototypical AN vectors in vector spaces can be used as an indicator for AN identification. The distributional study we conducted shows that in French a semantically coherent class of ANs includes nouns with various morphological profiles. ANs may be affixed, converted or compound nouns, as well as opaque or morphologically simple nouns. When derived, ANs do not stem necessarily from verbs: the morphological base can be a noun or an adjective, and it is not semantically restricted to the denotation of actions. In other words, ANs are not necessarily derived from lexemes that semantically involve an agent. Agentivity in the nominal domain is not necessarily imported (in particular from the verbal domain), but can build directly in the semantic structure of ANs.

Suffixes used to form ANs in French are diverse and include at least *-eur*, *-ier*, *-iste*, *-ien*, *-aire* and *-ard*. Some of these suffixes are more frequently agentive than others, but they all allow for a variety of interpretations. In addition to their agentive interpretation, *-eur* can denote instruments, *-ier* containers, *-aire* beneficiaries, *-ien* inhabitants, etc. There appears to be no exclusive agentive suffix, and no exclusively agentive suffix.

We also investigated the subclassification of ANs and argued that the traditional distinction between functional and occasional agents had to be supplemented with a third subclass, that of behavioral agents. The distinction between the three subclasses is supported by distributional data and extends to non-*-eur*-suffixed ANs, presumably in correlation with morphological specificities as regards word-formation, affix and base selection. The traditional view that the distinction between functional and occasional agents is not formally marked in contemporary languages like French hides a more complex reality. While *-eur* is a versatile agentive suffix, there seems to be predilections as to the type of agents denoted by nouns ending with other agentive suffixes. We have formulated several hypotheses in that respect that have to be tested in future work.

As far as methodology is concerned, our study shows that with a careful processing of the linguistic data—in the present case, an accurate selection of monosemous deverbal ANs—distributional semantics tools can help answering basic research questions and support fine-grained theoretical distinctions. Some further research on nominal agentivity driven by distributional semantics methods can be considered. The morphosemantic properties of the different agentive suffixes can be examined in a distributional study, to determine the exact conditions of affix rivalry in AN formation and possibly refine their semantic analysis. Affix polysemy should also be investigated in connection with agentive meaning subtypes, to evaluate the semantic

organization of affixes in the morphological system. Although we focused here on French ANs, the distributional analysis of ANs could be used in the study of other languages, or contrastively, as long as prototypical classes of ANs can be identified and used as a baseline to assess nominal agentivity.

Acknowledgements We are grateful to the two anonymous reviewers and the editor for their helpful comments on an earlier version of this paper. We also wish to thank Cécile Fabre, Nabil Hathout, Alizée Lombard and Sandra Schwab for valuable discussions. The distributional semantics models used in this study were computed on the Osirim platform administrated by the IRT computer science lab and supported by the National Center for Scientific Research (CNRS), the Région Midi Pyrénées, the French Government and the European Regional Development Fund (ERDF). The open access funding was provided by the University of Fribourg.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aleksandrova, A. (2013). *Noms humains de phase: problèmes de classifications ontologiques et linguistiques*. PhD thesis, University of Strasbourg, Strasbourg.
- Alexiadou, A., & Schäfer, F. (2006). Instrument subjects are agents or causers. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West coast conference on formal linguistics, cascadiilla proceedings project*, Somerville, MA (pp. 40–48).
- Alexiadou, A., & Schäfer, F. (2010). *On the syntax of episodic vs. dispositional -er nominals* (pp. 9–38). Berlin: Mouton de Gruyter.
- Anscombe, J. C. (2001). A propos des mécanismes sémantiques de formation de certains noms d'agent en français et en espagnol. *Langages*, 143, 28–48.
- Anscombe, J. C. (2003). L'agent ne fait pas le bonheur: agentivité et aspectualité dans certains noms d'agent en espagnol et en français. *Thélème, Revista Complutense de Estudios Franceses*, 11, 11–27.
- Antoniak, M., & Mimno, D. (2018). Evaluating the stability of embedding-based word similarities. *Transactions of the Association for Computational Linguistics*, 6, 107–119.
- Barker, C. (2008). Possessives and relational nouns. In C. von Stechow, K. Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning*, Berlin: Mouton de Gruyter.
- 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: Mouton de Gruyter.
- Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6, 213–234. <https://doi.org/10.1146/annurev-linguistics-011619-030303>.
- Bonami, O., & Paperno, D. (2018). Inflection vs. derivation in a distributional vector space. *Lingue & Linguaggio*, 17(2), 173–196.
- Booij, G. (1986). Form and meaning in morphology: The case of Dutch agent nouns. *Linguistics*, 24, 503–517.
- Booij, G., & Lieber, R. (2004). On the paradigmatic nature of affixal semantics in English and Dutch. *Linguistics*, 42, 327–358.

- 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.
- Castella, M. (2014). *Bare predicates: Between syntax and semantics*. PhD thesis, University of Verona, Verona.
- Cohen, A. (2016). A semantic explanation for the external argument generalization. *Morphology*, 26(1), 91–103.
- Cruse, D. A. (1973). Some thoughts on agentivity. *Journal of Linguistics*, 9, 11–23.
- De Swart, H., Winter, Y., & Zwarts, J. (2007). Bare nominals and reference to capacities. *Natural Language and Linguistic Theory*, 25(1), 195–222.
- DeLancey, S. (1984). Notes on agentivity and causation. *Studies in Languages*, 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*. Larousse: Paris.
- Erk, K., Padó, S., & Padó, U. (2010). A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics*, 36(4), 723–763.
- Fillmore, C. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). New York: Holt, Rinehart & Winston.
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In J. R. Firth (Ed.), *Studies in linguistic analysis* (pp. 1–32). Oxford: Basil Blackwell.
- 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* (pp. 177–196). Barcelona: IULA-universitat Pompey Fabra.
- Godard, D., & Jayez, J. (1993). Types nominaux et anaphores : le cas des objets et des événements. In W. De Mulder, L. Tasmowski-De Ryck, & C. Vetters (Eds.), *Anaphores temporelles et (in-)coherence, Cahiers chronos* (Vol. 1, pp. 41–58). Amsterdam: Rodopi.
- Grimm, S. (2011). Semantics of case. *Morphology*, 21(3–4), 515–544.
- Grimm, S. (2013). The bounds of subjecthood: Evidence from instruments. In *Proceedings of the 33rd meeting of the Berkeley linguistic society* (pp. 178–189).
- Gruber, J. S. (1967). Look and see. *Linguistics*, 43, 937–947.
- Halliday, M. A., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2–3), 146–162.
- Holisky, D. A. (1987). The case of the intransitive subject in Tsova-Tush (Batsbi). *Lingua*, 71(1–4), 103–132.
- Huyghe, R. (2015). Les typologies nominales : présentation. *Langue Française*, 1, 5–27.
- Huyghe, R. (to appear) La construction morphosémantique des noms d'agent en français. In: A. Aleksandrova & J. P. Meyer (Eds.), *Nommer l'humain : descriptions et catégorisations*. Paris: L'Harmattan.
- Huyghe, R., & Tribout, D. (2015). Noms d'agents et noms d'instruments : le cas des déverbaux en *-eur*. *Langue Française*, 185, 99–112.
- Ježek, E., & Varvara, R. (2015). Instrument subjects without instrument role. In *Proceedings of the 11th joint ACL-ISO workshop on interoperable semantic annotation (ISA-11)*.
- Kamp, H., & Rossdeutscher, A. (1994). Remarks on lexical structure and DRS-construction. *Theoretical Linguistics*, 20(2–3), 97–164.
- Keyser, S. J., & Roeper, T. (1984). On the middle and ergative constructions in English. *Linguistic Inquiry*, 15(3), 381–416.
- Kintsch, W. (2001). Predication. *Cognitive Science*, 25(2), 173–201.
- Kipper Schuler, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD thesis, University of Pennsylvania, Philadelphia.
- Lakoff, G. (1977). Linguistic gestalts. In *Papers from the thirteen regional meeting* (Vol. 13, pp. 236–287). Chicago: Chicago Linguistic Society, University of Chicago.
- Lapesa, G., Kawaletz, L., Plag, I., Andreou, M., Kisselew, M., & Padó, S. (2018). Disambiguation of newly derived nominalizations in context: A distributional semantics approach. *Word Structure*, 11(3), 277–312.
- Lenci, A. (2011). Composing and updating verb argument expectations: A distributional semantic model. In *Proceedings of the 2nd workshop on cognitive modeling and computational linguistics* (pp. 58–

- 66).
- Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, 4, 151–171.
- Lieber, R. (2016). *English nouns: The ecology of nominalization*. Cambridge: Cambridge University Press.
- Lieber, R., & Andreou, M. (2018). Aspect and modality in the interpretation of deverbal *-er* nominals in English. *Morphology*, 28(2), 187–217.
- Lowder, M. W., & Gordon, P. C. (2015). Natural forces as agents: Reconceptualizing the animate–inanimate distinction. *Cognition*, 136, 85–90.
- Mahlberg, M. (2005). *English general nouns: A corpus theoretical approach* (Vol. 20). Amsterdam: Benjamins.
- Matushansky, O., & Spector, B. (2005). Tinker, tailor, soldier, spy. In E. Maier, C. Bary, & J. Huitink (Eds.), *Proceedings of Sinn und Bedeutung* (Vol. 9, pp. 241–255). Nijmegen: NCS.
- Mickus, T., Bonami, O., & Paperno, D. (2019). Distributional effects of gender contrasts across categories. *Proceedings of the Society for Computation in Linguistics*, 2, 174–184.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. Preprint [arXiv:13013781](https://arxiv.org/abs/1301.3781).
- Mikolov, T., Yih, Wt. & 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* (Vol. 13, pp. 746–751).
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). *Lexique 2: A new French lexical database. Behavior Research Methods, Instruments, & Computers*, 36(3), 516–524.
- Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1), 71–106.
- Partee, B. H., & Borschev, V. (2003). Genitives, relational nouns, and argument-modifier ambiguity. In E. Lang, C. Maienborn, & C. Fabricius-Hansen (Eds.), *Modifying adjuncts* (Vol. 4, pp. 67–112). Berlin: Mouton de Gruyter.
- Pierrejean, B. (2020). *Qualitative evaluation of word embeddings: investigating the instability in neural-based models*. PhD thesis, University Toulouse Jean Jaurès, Toulouse.
- Pierrejean, B., & Tanguy, L. (2018). Towards qualitative word embeddings evaluation: Measuring neighbors variation. In *Conference of the North American chapter of the association for computational linguistics: Student research* (pp. 32–39). New Orleans: Workshop.
- Rapaport Hovav, M., & Levin, B. (1992). *-Er* nominals: Implications for the theory of argument structure. In T. Stowell & E. Wehrli (Eds.), *Syntax and semantics* (Vol. 26, pp. 127–153). New York: Academic Press.
- Roché, M. (2003). Catégorisation et recatégorisation en morphologie dérivationnelle : le cas de la dérivation en *-ier(e)*. In G. Coll & J. P. Régis (Eds.), *Travaux linguistique du CerLiCO : Vol. 16. Morphosyntaxe du lexique : catégorisation et mise en discours, Actes du Colloque de Tours, 7–8 juin 2002* (pp. 75–92). Rennes: Presses Universitaires de Rennes.
- 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 (Eds.), *Des unités morphologiques au lexique* (pp. 69–143). Paris: Hermès.
- 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)*. PhD thesis, University of Stockholm, 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.
- Roy, I., & Soare, E. (2014). Les noms d'humains dérivés de participes : nominalisations en *-ant* et *-e/ll/u*. In *SHS web of conferences* (Vol. 8, pp. 3197–3208). Les Ulis: EDP Sciences
- Ryder, M. E. (1999). Bankers and blue-chippers: An account of *-er* formations in present-day English. *English Language and Linguistics*, 3(2), 269–297.
- Schlesinger, I. M. (1989). Instruments as agents: On the nature of semantic relations. *Journal of Linguistics*, 25(1), 189–210.
- Schnedecker, C., & Aleksandrova, A. (2016). Les noms d'humains en *-aire* : essai de classification. In F. Neveu, G. Bergounioux, M.-H. Côté, J.M.Fournier, L. Hriba, & S. Prévost (Eds.), *Congrès mondial de linguistique française 2016*, Paris: Institut de Linguistique Française.
- Sleeman, P., & Verheugd, E. (2004). Action and agent nouns in French and polysemy. In D. Willems, B. Defranq, T. Coleman, & D. Noël (Eds.), *Contrastive analysis in language: Identifying linguistic units of comparison* (pp. 137–154). New York: Palgrave Macmillan.

- Tribout, D. (2010). *les conversions de nom à verbe et de verbe à nom en français*. PhD thesis, University of Paris Diderot, Paris.
- Tribout, D., Barque, L., Haas, P., & Huyghe, R. (2014). De la simplicité en morphologie. In F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefer, & S. Prevost (Eds.), *SHS web of conferences* (Vol. 8, pp. 1879–1890). Les Ulis: EDP Sciences
- Urieli, A. (2013). *Robust French syntax analysis: reconciling statistical methods and linguistic knowledge in the Talismane toolkit*. PhD thesis, University of Toulouse Jean Jaurès, Toulouse.
- Van Peteghem, M. (1993). *La détermination de l'attribut nominal: étude comparative de quatre langues romanes (français, espagnol, italien, roumain)*. Brussels: Brepols.
- Van Valin, R., & LaPolla, R. (1997). *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
- Van Valin, R. D., & Wilkins, D. (1996). The case for 'effector': Case roles, agents, and agency revisited. In M. Shibatani & S. A. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 289–322). Oxford: Oxford University Press.
- Varvara, R., Lapesa, G., & Padó, S. (2016). Quantifying regularity in morphological processes: An ongoing study on nominalization in German. In *ESSLLI DSALT workshop: Distributional semantics and semantic theory*.
- Vikner, C., & Jensen, P. A. (2002). A semantic analysis of the English genitive: Interaction of lexical and formal semantics. *Studia Linguistica*, 56(2), 191–226.
- Villoing, F. (2009). Les mots composés VN. In B. Fradin, F. Kerleroux, & M. Plénat (Eds.), *Aperçus de morphologie du français* (pp. 175–198). Paris: Presses Universitaires de Vincennes.
- Wauquier, M., Fabre, C., & Hathout, N. (2018). Différenciation sémantique de dérivés morphologiques à l'aide de critères distributionnels. In F. Neveu, B. Harmegnies, L. Hriba, & S. Prevost (Eds.), *SHS web of conferences* (Vol. 46, pp. 1156–1170). Les Ulis: EDP Sciences.
- 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, pp. 111–121).
- Winther, A. (1975). Note sur les formations déverbiales en *-eur* et *-ant*. *Cahiers de Lexicologie*, 26, 35–54.