Agent nouns and affix selection

Towards a morphosemantic typology of French agent nouns based on distributional semantics

Richard Huyghe [‡]
Marine Wauquier [✳]

[‡] Université de Fribourg
[✳] CLLE, CNRS & Université Toulouse Jean Jaurès

Extension of Agent nouns

- The class of Agent nouns [ANs] has been given different extensions in the literature:
  - Nouns derived from verbs, e.g. French *-eur* nouns (Benveniste 1975, Winther 1975, Ulland 1993, Anscombre 2001, Fradin & Kerleroux 2003, Sleeman & Verheugd 2004, Roy & Soare 2012, Huyghe & Tribout 2015), VN compounds (Ulland 1993, Villoing 2003, 2009, Rosenberg 2008)
  - Nouns derived not necessarily from verbs, e.g. French *-aire*, *-ien*, *-ier*, *-iste*, *-logue* nouns (Dubois 1962, Roché 2004, 2011, Namer & Villoing 2014, Cartoni *et al.* 2015, Schnedecker & Aleksandrova 2016)
  - Underived nouns (Blinkenberg 1960, Huyghe to appear)
- The agentitivity of some nouns (e.g. *-ant* nouns) is overtly discussed (Winther 1975, Anscombre 2003, Roy & Soare 2012 *contra* Lerat 1984, Rosenberg 2008)

Research questions

- Which nouns do form a semantically coherent class of ANs?
- Does the categorization as AN depend on (i) the morphological properties of nouns, (ii) the PoS or (iii) the semantic type of the base?
- If different morphological processes yield to ANs, do they correlate with semantic specificities?

Distributional semantics and ANs

- Can the distributional semantics tools developed in computational linguistics help in answering research questions?
- In a first study based on distributional semantics, we showed that various derivational processes were involved in the creation of French ANs
- We now intend to compare the distributional profiles of morphologically diverse ANs, in order to investigate their potential morphosemantic specificities

Plan

1. Identifying Agent nouns in French
2. Primary target Agent nouns
3. Centroids: results and analysis
4. Clusters: results and analysis
5. Conclusion

Identifying Agent nouns in French

Defining agentivity

- Agent is often considered a difficult notion to define (Cruse 1973, DeLancey 1984, Schlesinger 1989, Dowty 1991, *i.a.*). Debates focus on:
  - The intentionality of Agents
  - The animacy of Agents
  - The dynamicity of the predicates involved
- We define Agents as effectuators (i.e. entities deploying energy to carry out actions) that are prototypically animate and intentional
- At the lexical level, nouns can be categorized as ANs if they denote their referent as the Agent of an intrinsically specified action

Action component

- The categorization as AN needs to guarantee the existence of an action in the semantic structure of a noun
- **Possible resolution**: ANs are deverbal nouns that denote the agentive argument of the base verb
- **Difficulty**: Morphologically heterogeneous but semantically homogeneous classes are split
  1. SCULPTEUR 'sculptor', GUÉRISSEUR 'healer', RÉDACTEUR 'writer', MANIFESTANT 'demonstrator', PROTESTATAIRE 'protestor'
  2. ARTISTE 'artist', MÉDECIN 'doctor', SCRIBE 'scribe', GRÉVISTE 'striker', ÉMEUTIER 'rioter'
- Can non-deverbal nouns be considered ANs? If so, how can we highlight their agentive meaning and delineate the class of ANs?
- We use a distributional approach to answer these questions.

Distributional semantics

- The distributional hypothesis correlates the semantic similarity of words with the amount of context they share (Harris 1954, Firth 1957)
  - On the basis of context sharing in corpora, the meaning of words are represented as vectors in a multidimensional space (DSM)
  - The geometrical nature of those representations allows for their manipulation: distance calculation, addition, average, etc.
- We assume that a unified representation of an entire nominal class can be computed by averaging the vectors of its members (Kintsch 2001, Wauquier *et al.* 2018, Bonami & Paperno 2018)
  - The nearest neighbours of the average vector (**centroid**) are considered as representative instances of the nominal class

Experimental set-up

- We build our DSMs from the French Wikipedia corpus preprocessed with Talismane (Urieli & Tanguy 2013) and used the default settings of Word2Vec (Mikolov et al 2013)
    - CBOW, frequency threshold of 5, 5-sized context window
- We work on the basis of the results from 5 distinct DSMs built with the same settings so as to control the variation induced by stochastic methods involved in the DSMs training (Pierrejean and Tanguy, 2018)

Selecting prototypical Agent nouns

- We first extract *-eur*/*-euse*/*-rice* nouns from the lexique.org lexical resource to constitute a prototypical set of ANs
- 2215 nouns are listed. Monosemous deverbal ANs are manually selected (Huyghe & Tribout 2015)
- 1121 ANs are listed, out of which 558 occur in the Wikipedia corpus with a frequency ≥ 5
  - The selected nouns are almost exclusively human-denoting nouns [HNs]
- Vectors are built for all these nouns; so is their centroid
- The nearest neighbours to the centroid are scrutinized

Neighbourhood of the agentive N-*eur* centroid

Noun	Proximity
PLOMBIER	0.770
TRUAND	0.750
ESCROC	0.748
RABATTEUR	0.730
COIFFEUR	0.726
PRESTIDIGITATEUR	0.719
PROXÉNÈTE	0.715
GARAGISTE	0.709
GANGSTER	0.706
CAMIONNEUR	0.704
MALFRAT	0.701
PICKPOCKET	0.700
CUISINIER	0.699
VOLEUR	0.694
VOYOU	0.691
SOIGNEUR	0.688
MAGICIEN	0.686
CHARLATAN	0.686
RECELEUR	0.686
DOMPTEUR	0.685
FARCEUR	0.683
CAMBRIOLEUR	0.678
COLPORTEUR	0.675
DÉMÉNAGEUR	0.674
JARDINIER	0.674
BIJOUTIER	0.674
CUISTOT	0.671
DOMESTIQUE	0.669
DROGUÉ	0.667
…	

Analysis of the 100 nearest neighbours

- 29% of the 100 nearest neighbours are primary targets [PTs], i.e. nouns used in the creation of the centroid
- All 100 neighbours are HNs, with the exception of CHIEN 'dog'
- Most neighbours are suffixed nouns but the list also includes morphologically simple nouns (PROXÉNÈTE 'pimp'), converse nouns (DROGUÉ 'drug addict'), underived complex nouns (APPRENTI 'apprentice'), etc.
- Different suffixes are used (*-eur*, *-ier*, *-iste*, *-ard*, etc.)
- The bases are syntactically and semantically diverse (verbs, adjectives, action-, object-, property-, domain-denoting nouns, etc.)

Analysis of the 100 nearest neighbours

	Suff	Spl	Cpx	Conv	Cpd	Undet	Extra
-EUR	62	13	10	6	5	3	1

Table 1: Morphology of the 100 nearest neighbours

	-eur	*ø*	*-ier*	*-iste*	*-ard*	*-on*	other
-EUR	32	27	18	9	3	3	8

Table 2: Affixation of the 100 nearest neighbours

Analysis of derived nouns among the 100 nearest neighbours

	Verb	Noun	Adjective
-EUR	37	28	3

Table 3: PoS of the base

	Action	Object	Domain	Property	Other
-EUR	44	13	6	3	2

Table 4: Semantic type of the base

Semantic information

- PTs used to build the *-eur* centroid have 2 fundamental semantic features: (i) human and (ii) agentive
- To determine whether the centroid is sensitive to both features, we calculate its proximity to non-agentive HNs:
  - General HNs (PERSONNE 'person', GENS 'people')
  - Phase HNs (ADOLESCENT 'teenager', VIEILLARD 'old man')
  - Relational HNs (FILS 'son', OTAGE 'hostage')
  - Demonyms (FIDJIEN 'Fijian', GENEVOIS 'Genevan')
- The scores are (much) lower than in the case of the first 200 neighbours of the *-eur* centroid
- We hypothesize that the proximity to the *-eur* centroid reveals a combination of (i) & (ii), and can thus be used as a test for nominal agentive meaning

Analysis of AN candidates

- To evaluate the agentivity of AN candidates, we compare the proximity scores between their vectors and different HN centroids. Several types of candidates are tested:
  - Denominal -*eur* nouns
  - -*aire* nouns
  - -*iste* nouns
  - -*ien* nouns
  - -*er*/-*ier* nouns
  - -*ant* nouns
  - Various converse nouns
  - Morphologically simple nouns

A sample of AN candidates

Noun	Agent	Phase	Relation	Inhab
FARCEUR 'joker'	0.683	0.641	0.373	0.195
FAUSSAIRE 'forger'	0.656	0.428	0.289	0.277
PIANISTE 'pianist'	0.430	0.214	0.398	0.114
PHARMACIEN 'pharmacist'	0.564	0.413	0.403	0.263
HORLOGER 'clockmaker'	0.543	0.351	0.360	0.325
ATTAQUANT 'attacker'	0.501	0.268	0.268	0.173
GUIDE 'guide'	0.379	0.240	0.151	0.382
MÉDECIN 'doctor'	0.605	0.490	0.483	0.280

Table 5: Proximity of AN candidates with HNs centroids

Summary

- Nominal agentivity is (at least to some extent) captured by distributional analysis
- The proximity to the *-eur* AN centroid can be an indicator for AN detection
- A semantically coherent class of ANs includes nouns with various morphological profiles
- When ANs are derived, the base is not necessarily a verb
- ANs are not necessarily derived from words that semantically involve an Agent

Primary target Agent nouns

Morphological diversity

- We question the existence of different derivational processes for ANs: does the variation of morphological construction correlate with semantic specificities among ANs?
- To answer that question, we compare the distributional profiles of *-aire*, *-ant*, *-eur*, *-ien*, *-ier*, *-iste* ANs
- AN candidates are extracted from lexique.org and manually selected

Selection of AN candidates

- For each affix taken into account, we exclude from the list:
  - Nouns unanalysable as synchronically derived (AUTISTE 'autistic')
  - Nouns with at least one inanimate sense (CERISIER 'cherry tree')
  - Nouns with at least one general, phasal, relational or demonymic sense (QUADRAGÉNAIRE 'quadragenarian', JURASSIEN 'Jurassian')
  - Nouns with at least one axiological sense (TROTSKISTE 'Trotskyist')
  - Nouns with at least one sense denoting a non-agentive (i.e. experiencer, beneficiary, etc.) argument of the base (CROYANT 'believer', CESSIONNAIRE 'assignee')

Controlling primary targets

- AN candidates are controlled as regards their proximity to the *-eur* centroid (in contrast to lower proximity scores to other HN centroids)
- Are finally selected as PTs:
  - 28 *-aire* ANs (DIAMANTAIRE 'diamond merchant', PLAGIAIRE 'plagiarist', MANDATAIRE 'fiduciary')
  - 27 *-ant* ANs (COMMUNICANT 'communicator', FABRICANT 'manufacturer', PLAIGNANT 'complainant')
  - 577 *-eur* ANs (COIFFEUR 'hairdresser', FARCEUR 'prankster', ROCKEUR 'rocker')
  - 31 *-ien* ANs (GRAMMAIRIEN 'grammarian', OPTICIEN 'optician', TECHNICIEN 'technician')
  - 107 *-ier* ANs (COUTURIER 'fashion designer', GUICHETIER 'ticket seller', MEURTRIER 'murderer')
  - 146 *-iste* ANs (BASSISTE 'bassist', ESSAYISTE 'essayist', ORTHOPHONISTE 'speech therapist')

PT analysis: PoS of the base

Affix	Verb	Noun	Adjective
-AIRE	2	26	0
-ANT	27	0	0
-EUR	558	19	0
-IEN	0	29	2
-IER	4	103	0
-ISTE	3	134	9

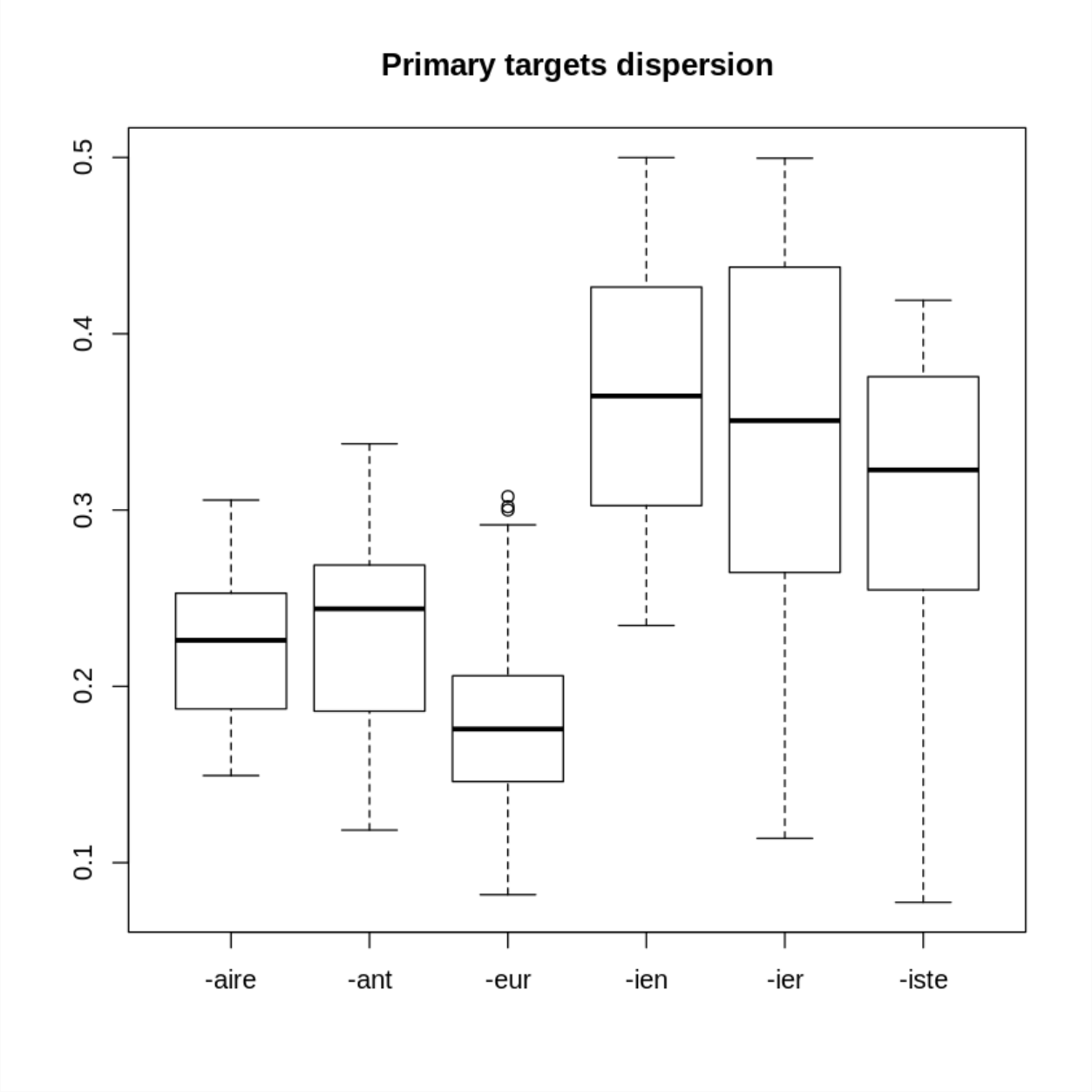
Table 6: PoS of the base

PT analysis: semantic type of the base

Affix	Act	Obj	Ppt	Dom	Cog	Obj*Cog	Other
-AIRE	13	4	3	2	1	2	2
-ANT	27	0	0	0	0	0	0
-EUR	559	3	1	13	0	0	1
-IEN	1	1	0	25	3	0	1
-IER	9	85	0	3	1	3	6
-ISTE	15	37	11	57	2	13	12

Table 7: Semantic type of the base

PT analysis: dispersion



Centroids: results and analysis

Building centroids

- We build centroids for each set of ANs (following the methodology presented in section 1)
  - Proximity scores are calculated over 5 different models
- For each affix, the 50 nearest neighbours shared in the 5 models are analyzed

PT coverage

- Among the 50 nearest neighbours of each centroid:
  - In the case of -*aire*, 4% are PTs (i.e. 7% of -*aire* PTs)
  - In the case of -*ant*, 12% are PTs (i.e. 22% of -*ant* PTs)
  - In the case of -*eur*, 24% are PTs (i.e. 2% of -*eur* PTs)
  - In the case of -*ien*, 12% are PTs (i.e. 19% of -*ien* PTs)
  - In the case of -*ier*, 50% are PTs (i.e. 23% of -*ier* PTs)
  - In the case of -*iste*, 32% are PTs (i.e. 11% of -*iste* PTs)

Neighbourhood of the *-aire* centroid

Noun	Proximity
COMPTABLE	0.779
MÉDECIN	0.756
FONCTIONNAIRE	0.746
INFORMATEUR	0.721
COMMERÇANT	0.718
PHARMACIEN	0.717
MILITAIRE	0.713
EMPLOYÉ	0.712
FINANCIER	0.710
COMMIS	0.708
VÉTÉRINAIRE	0.703
COLLABORATEUR	0.698
OFFICIER	0.693
ENTREPRENEUR	0.692
MAGISTRAT	0.687
JURISTE	0.685
BANQUIER	0.684
ARMURIER	0.680
BIBLIOTHÉCAIRE	0.677
OUVRIER	0.671
SAVANT	0.670
LOBBYISTE	0.669
AVOCAT	0.665
RESPONSABLE	0.665
PRATICIEN	0.665
INDUSTRIEL	0.663
ADMINISTRATEUR	0.661
CITOYEN	0.660
TERRASSIER	0.659
…	

Neighbourhood of the *-ant* centroid

Noun	Proximity
DIRIGEANT	0.758
EMPLOYÉ	0.756
RESPONSABLE	0.731
INFORMATEUR	0.708
COLLABORATEUR	0.705
LOBBYISTE	0.700
ÉTUDIANT	0.694
EMPLOYEUR	0.693
MILITANT	0.693
TRAVAILLEUR	0.691
INTERVENANT	0.691
ACTIVISTE	0.690
ANALYSTE	0.690
SYMPATHISANT	0.688
DÉTENU	0.687
CITOYEN	0.674
AGENT	0.672
RETRAITÉ	0.671
INVESTISSEUR	0.671
DÉCIDEUR	0.670
EXPERT	0.669
TERRORISTE	0.669
CLIENT	0.667
MILITAIRE	0.667
COMPTABLE	0.665
DÉLINQUANT	0.660
PRATICIEN	0.660
TRANSFUGE	0.657
ENSEIGNANT	0.656
…	

Neighbourhood of the *-ien* centroid

Noun	Proximity
PHYSICIEN	0.840
STATISTICIEN	0.838
NEUROLOGUE	0.825
INFORMATICIEN	0.825
LINGUISTE	0.815
MATHÉMATICIEN	0.814
CHIMISTE	0.812
CARDIOLOGUE	0.808
BIOLOGISTE	0.807
PHYSIOLOGISTE	0.804
PSYCHIATRE	0.803
PÉDIATRE	0.802
JURISTE	0.801
NEUROCHIRURGIEN	0.801
OPHTALMOLOGUE	0.800
ANTHROPOLOGUE	0.800
SOCIOLOGUE	0.794
CRIMINOLOGUE	0.793
ÉCONOMISTE	0.792
PSYCHOLOGUE	0.790
HÉMATOLOGUE	0.790
HÉLLÉNISTE	0.782
ÉLECTROTECHNICIEN	0.782
GÉOPHYSICIEN	0.778
ETHNOLOGUE	0.778
GERMANISTE	0.777
PÉDAGOGUE	0.775
UROLOGUE	0.773
MÉDECIN	0.773
…	

Neighbourhood of the *-ier* centroid

Noun	Proximity
ARMURIER	0.862
CORDONNIER	0.860
BOULANGER	0.822
BIJOUTIER	0.817
MARÉCHAL-FERRANT	0.816
SERRURIER	0.812
CHARPENTIER	0.808
PLOMBIER	0.802
FORGERON	0.802
PERRUQUIER	0.797
ÉPICIER	0.788
BOTTIER	0.787
CHAUDRONNIER	0.786
BOUCHER	0.784
TANNEUR	0.783
MENUISIER	0.783
CHARRETIER	0.779
GARAGISTE	0.778
COIFFEUR	0.778
JARDINIER	0.777
AUBERGISTE	0.776
CAFETIER	0.774
QUINCAILLIER	0.772
PORTEFAIX	0.770
CONTREMAÎTRE	0.768
TAILLEUR	0.767
CUISINIER	0.767
TONNELIER	0.760
FRIPIER	0.759
…	

Neighbourhood of the *-iste* centroid

Noun	Proximity
VIOLONISTE	0.800
CLARINETTISTE	0.798
PHOTOGRAPHE	0.794
PÉDAGOGUE	0.784
OPHTALMOLOGUE	0.782
VIOLONCELLISTE	0.777
PIANISTE	0.775
MUSICOLOGUE	0.768
MUSICIEN	0.767
ÉCRIVAIN	0.764
FLÛTISTE	0.764
PRESTIDIGITATEUR	0.763
COMPOSITEUR	0.762
NATUROPATHE	0.759
NEUROLOGUE	0.759
PHONÉTICIEN	0.756
PÉDIATRE	0.755
CÉRAMISTE	0.753
MYCOLOGUE	0.752
ALTISTE	0.752
ETHNOMUSICOLOGUE	0.752
ORNITHOLOGUE	0.752
GERMANISTE	0.752
LINGUISTE	0.748
PSYCHIATRE	0.747
CLAVECINISTE	0.747
HAUTBOÏSTE	0.746
TROMPETTISTE	0.743
NEUROCHIRURGIEN	0.742
…	

First observations

- The neighbourhood of the *-iste* centroid seems to consist mainly of nouns which denote domain experts and persons who engage in domain-related activities
- The neighbourhood of the *-ien* centroid seems semantically close to that of *-iste* (with the exception of artist-denoting nouns which seem to be closer to *-iste*); *-logue* nouns appear to be privileged neighbours of *-iste* and *-ien*
- The neighbourhood of the *-ier* centroid seems mostly composed of 'manual' professions (craftspeople, storekeepers)

First observations

- The neighbourhood of the *-aire* centroid seems more heterogeneous; some neighbours denote office jobs; some episodic agents are included (COLLABORATEUR 'collaborator', CRIMINEL 'criminal', COMMISSIONNAIRE 'commissioner')
- The neighbourhood of the *-ant* centroid is also heterogeneous, with less profession nouns and more episodic agents (INTERVENANT 'speaker', COLLABORATEUR 'collaborator', TRANSFUGE 'defector'), which might correlate with the description of 'negative' actions (TERRORISTE 'terrorist', CRIMINEL 'criminal', DÉLINQUANT 'offender')

Analysis of the 50 nearest neighbours

Affix	*-aire*	*-ant*	*-eur*	*-ien*	*-ier*	*-iste*	*-logue*	*ø*	autre
-AIRE	5	3	7	5	12	3	0	7	8
-ANT	4	10	9	3	2	4	0	8	10
-EUR	1	0	15	1	9	5	0	13	-
-IEN	0	0	0	13	0	14	16	2	5
-IER	0	0	6	0	35	3	0	5	1
-ISTE	1	0	4	4	0	20	11	5	5

Table 8: Affixation of the 50 nearest neighbours

Analysis of the 50 nearest neighbours

Affix	Act	Obj	Ppt	Dom	Obj*Cog	Other
-AIRE	16	5	4	6	1	6
-ANT	23	1	8	2	0	4
-EUR	18	10	2	4	0	1
-IEN	0	0	2	24	0	0
-IER	10	29	0	2	0	0
-ISTE	3	1	2	20	0	2

Table 9: Semantic type of the base of the 50 nearest neighbours

Shared neighbours

	-AIRE	-ANT	EUR	-IEN	-IER	-ISTE
-AIRE	-	15	5	3	4	0
-ANT	15	-	1	0	0	0
-EUR	5	1	-	0	9	2
-IEN	3	0	0	-	0	13
-IER	4	0	9	0	-	0
-ISTE	0	0	2	13	0	-

Table 10: Shared neighbours between centroids

Average proximity between centroids

	-AIRE	-ANT	-EUR	-IEN	-IER	-ISTE
-AIRE	-	0,836	0,781	0,726	0,790	0,670
-ANT	0,836	-	0,780	0,612	0,620	0,570
-EUR	0,781	0,780	-	0,723	0,844	0,761
-IEN	0,726	0,612	0,723	-	0,670	0,913
-IER	0,790	0,620	0,844	0,670	-	0,710
-ISTE	0,670	0,570	0,761	0,913	0,710	-

Table 11: Average proximity score between centroids over 5 models

Analysis

- In the case of *-iste* and *-ien*, there is a congruence between:
  - The rather high proximity between the PTs of each affix
  - The semantic homogeneity of the neighbours of each centroid
  - The high number of shared neighbours between the 2 centroids
  - The high proximity between the 2 centroids
- *-iste* and *-ien* ANs have similar profiles. They denote domain experts and agents of intellectual activities (with a specificity for artists in the case of *-iste*)

Analysis

- In the case of *-aire* and *-ant*, there is a congruence between:
  - The semantic type of the base (Action)
  - The episodic feature in some PTs and neighbours, and the related tendency not to exclusively denote professions or hobbys
  - The rather important distance between PTs, corresponding to a greater semantic diversity of the nouns
  - The high number of shared neighbours between the 2 centroids
  - The high proximity between the 2 centroids
- *-aire* and *-ant* ANs are semantically more versatile than other ANs. That versatility is related to the predilection for action-denoting bases. This tendency is more prominent for *-ant* ANs

Analysis

- In the case of *-ier*, there is a congruence between:
  - The rather high proximity between PTs
  - The high number of shared PTs and neighbours
  - The morphological homogeneity of the neighbours
  - A predilection for object-denoting bases
- *-ier* ANs differ from other ANs in that they mostly denote professions in crafts or commerce

Analysis

- In the case of *-eur*, there is a congruence between:
  - The high dispersion of PTs
  - The PoS of the base (verbs)
  - The semantic type of the base (Action)
  - The morphosemantic heterogeneity of the neighbours
- *-eur* ANs are the most heterogeneous ANs. They can denote episodic, dispositional or institutional agents. Any kind of profession can be denoted as well

Clusters: results and analysis

Clustering of Agent Nouns

- In order to test or refine the semantic distinctions sketched in the analysis of centroids, we apply a clustering algorithm to PT vectors (i.e. a method to group vectors wrt their proximity in DSMs)
- Do these clusters match with ANs' morphological subclasses and correspond to the distinctions highlighted in the affix analysis?
- Is the class of ANs internally organized by semantic distinctions that do not correlate with morphological diversity?

Cluster analysis: affixes

Cluster	-AIRE	-ANT	-EUR	-IEN	-IER	-ISTE
c1	12	14	146	3	11	12
c2	9	7	129	6	79	27
c3	1	1	201	3	9	6
c4	6	5	101	19	8	101

Table 12: Affix distribution in 4 clusters

Cluster analysis: affixes

Cluster	-AIRE	-ANT	-EUR	-IEN	-IER	-ISTE
c1	1	0	33	15	3	82
c2	6	6	40	3	22	12
c3	3	2	116	1	10	8
c4	0	3	53	2	5	11
c5	0	0	53	0	2	2
c6	1	0	57	0	2	0
c7	3	1	66	2	52	12
c8	7	3	60	2	0	10
c9	0	0	37	5	5	5
c10	7	11	62	1	6	3

Table 13: Affix distribution in 10 clusters

Cluster analysis: semantic type of the base

Cluster	Act	Obj	Ppt	Dom	Cog	Obj*Cog	Other
c1	176	4	3	4	3	2	5
c2	143	90	0	12	1	2	9
c3	201	10	5	4	0	0	0
c4	106	24	19	67	3	14	7

Table 14: Semantic type of the base (4 clusters)

Cluster analysis: semantic type of the base

Cluster	Act	Obj	Ppt	Dom	Cog	Obj*Cog	Other
c1	34	17	8	59	1	10	5
c2	52	25	2	5	1	0	4
c3	126	4	2	2	1	2	2
c4	54	7	6	5	1	0	1
c5	53	3	1	0	0	0	1
c6	57	3	0	0	0	0	0
c7	68	60	0	3	0	2	3
c8	69	0	2	5	2	2	2
c9	34	6	6	5	1	0	1
c10	79	3	0	3	0	2	2

Table 15: Semantic type of the base (10 clusters)

Summary

- The clustering method confirms some specificities of agentive affixes (e.g. the versatility of *-eur* as opposed to the homogeneity of *-iste*, *-ier*, *-ien*; the convergence between *-iste* and *-ien*, and between *-aire* and *-ant*)
- Clusters to some extent depend on the semantic type of the base (e.g. ANs derived from domain-denoting bases and from object-denoting bases are quite dense)
- The highly variable and sometimes quite small number of PTs may restrain the conclusions drawn out of clustering

Conclusion

Morphological heterogeneity

- A class of ANs based on distributional properties consists of morphologically heterogeneous nouns
- ANs may be derived or underived nouns
- When derived, ANs stem mostly from verbs and nouns. The semantic type of the base is not restricted
- Some affixes are prominently agentive, but there is no exclusive agentive affix, and no exclusively agentive affix

Semantic diversity

- Semantic specificities emerge for the different agentive affixes
- These specificities correlate (i) the affix, (ii) the grammatical and semantic selection of the base, (iii) the Agent type
- The semantic and distributional homogeneity of ANs depends on the affix
- The semantic proximity between ANs depends on the affix

Future research

- What are the semantic properties of non-surveyed ANs (converse nouns, -*logue* nouns, -*graphe* nouns, etc.)?
- Can the polysemy of the affixes involved in the formation of ANs shed light on their agentive specificities?
- Are distributional similarities among ANs defined by semantic properties other than those correlated to morphology?

References i

- Anscombre, Jean-Claude. 2001. À propos des mécanismes sémantiques de formation de certains noms d'agent en français et en espagnol. *Langages* 143, 28-48.
- Anscombre, Jean-Claude. 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*, 11-27.
- Benveniste, Emile. 1975. Noms d'agent et noms d'action en indo-européen. Paris : Maisonneuve.
- Bonami, Olivier & Denis Paperno. 2018. Inflection vs. derivation in a distributional vector space. *Lingue e Linguaggio* 2/2018, 173-196.
- Cartoni, B., Namer, F. & Lignon, S. A Cross-Linguistic Insight on Agentive Noun Formation in Italian and French. *Selected Papers from the 8th Décembrettes : Morphology in Bordeaux, Carnets de Grammaire*, vol. 22 : 81-98.
- Cruse, David Alan. 1973. Some thoughts on agentivity. *Journal of Linguistics* 9, 11-23.
- DeLancey, Scott. 1984. Notes on agentivity and causation. *Studies in Language* 82, 181-213.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67/3, 547-619.
- Dubois, Jean. 1962. *Etude sur la dérivation suffixale en français moderne et contemporain*. Paris : Larousse.
- Fradin, Bernard & Françoise Kerleroux. 2003. Quelle base pour les procédés de la morphologie constructionnelle ? *Silexicales* 3, 76-84.
- Huyghe, Richard & Delphine Tribout. 2015. Noms d'agents et noms d'instruments : le cas des déverbaux en *-eur*. *Langue Française* 185, 99-112.
- Kintsch, Walter. 2001. Predication. *Cognitive science*, 25(2), 173-202.
- Lerat, Pierre. 1984. Grammaire des noms d'agent en -ant en français contemporain. *Cahiers de Lexicologie* 24, 23-29.
- Namer, Fiammetta & Florence Villoing. 2014. Sens morphologiquement construit et procédés concurrents : les noms de spécialistes en –logue et –logiste. *Revue de Sémantique et de Pragmatique* 35-36, 7-26.
- Pierrejean, Bénédicte & Tanguy Ludovic. 2018. Étude de la reproductibilité des word embeddings : repéragedes zones stables et instables dans le lexique. In *Traitement Automatique des Langues Naturelles (TALN)*.

References ii

- Roché, Michel. 2004. Mot construit ? Mot non construit ? Quelques réflexions à partir des dérivés en -ier(e). *Verbum*, 26, 459-480.
- Roché, Michel. 2011. Quel traitement unifié pour les dérivations en -isme et en -iste? In *Des unités morphologiques au lexique*, M. Roché, G. Boyé, N. Hathout, S. Lignon, M. Plénat (éds), Paris: Hermès, 69-143.
- Rosenberg, Maria. (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).* Thèse de doctorat. Université de Stockholm.
- Roy, Isabelle & Elena Soare. 2012. L'enquêteur, le surveillant et le détenu : les noms déverbaux de participants aux événements, lectures événementielles et structure argumentale. *Lexique* 20, 207-231.
- Schlesinger, I. M. 1989. Instruments as agents: on the nature of semantic relations. *Journal of Linguistics*, 25-1, 189-210.
- Schnedecker, Catherine & Angelina Aleksandrova. 2016. Les noms d'humains en -aire : essai de classification. In *Congrès Mondial de Linguistique Française 201*6, Neveu, F., Bergounioux, G., Côté, M.-H., Fournier, J.-M., Hriba, L., Prévost, S. (éds), Paris : Institut de Linguistique Française.
- Sleeman, Petra & Els Verheugd. 2004. Action and agent nouns in French and polysemy. In *Contrastive Analysis in Language: Identifying Linguistic Units of Comparison*, D. Willems, B. Defranq, T. Colleman, D. Noël (eds.), New York: Palgrave McMillan, 137-154.
- Ulland, Harald. 1993. *Les nominalisations agentive et instrumentale en français moderne*. Berne: Peter Lang.
- Urieli, Assaf Tanguy, Ludovic (2013). L'apport du faisceau dans l'analyse syntaxique en dépendances par transitions : études de cas avec l'analyseur Talismane. In *Actes de la conférence Traitement Automatique des Langues Naturelles (TALN 2013).* Les Sables d'Olonne, France.
- Villoing, Florence. 2003. Les mots composés VN du français: arguments en faveur d'une construction morphologique. Cahiers de Grammaire 28, 183-196.
- Villoing, Florence. 2009. Les mots composés VN. In *Aperçus de morphologie du français*, B. Fradin, F. Kerleroux, M. Plénat, M. (éds), Paris : Presses Universitaires de Vincennes, 175-198.
- Wauquier, Marine ; Cécile Fabre & Nabil Hathout. 2018. Différenciation sémantique de dérivés morphologiques à l'aide de critères distributionnels. In *Congrès Mondial de Linguistique Française 2018*, F. Neveu, B. Harmegnies, L. Hriba, S. Prévost (éds), Paris : Institut de Linguistique Française.
- Winther, André. 1975. Note sur les formations déverbales en *-eur* et en *-ant*. *Cahiers de lexicologie* 26, 35-54.