

Les noms d'agent en français : Une approche distributionnelle

Richard Huyghe ‡
Marine Wauquier †

‡ Université de Fribourg

† LLF, CNRS & Université de Paris

La classe des noms d'agent en question

- La catégorie de 'nom d'agent' [Nag] est employée dans la littérature avec des extensions diverses :
 - N construits à partir de V, ex. N-*eur* (Benveniste 1948, Winther 1975, Ulland 1993, Anscombe 2001, Fradin & Kerleroux 2003, Sleeman & Verheugd 2004, Huyghe & Tribout 2015), composés VN (Ulland 1993, Rosenberg 2008, Villoing 2009)
 - N construits non nécessairement à partir de V, ex. N-*aire*, N-*ien*, N-*ier*, N-*iste*, N-*logue* (Dubois 1962, Roché 2004, 2011, Namer & Villoing 2014, Cartoni *et al.* 2015, Schnedecker & Aleksandrova 2016)
 - N non nécessairement construits (Blinkenberg 1960, Huyghe à paraître)
- NB : L'agentivité de certains cas est expressément discutée, ex. N-*ant* (Anscombe 2003, Roy & Soare 2012 *contra* Lerat 1984, Rosenberg 2008).

La classe des noms d'agent en question

- De quels N une classe de Nag sémantiquement cohérente est-elle constituée?
- L'appartenance à une telle classe dépend-elle des propriétés morphologiques des N, en particulier du procédé formel de construction des N, de la catégorie grammaticale ou du type sémantique des éventuelles bases de construction?
- En cas de coexistence de plusieurs procédés de construction des Nag, y a-t-il des propriétés distinctives de chaque procédé? Si oui, quelles sont-elles?

Le recours à la sémantique distributionnelle

- Pour répondre à ces questions, les outils de la sémantique distributionnelle développés en linguistique informatique peuvent-ils être utiles?
 - L'agentivité nominale peut-elle être mise en évidence dans l'analyse distributionnelle?
 - Comment isoler et accéder au contenu sémantique qui fonde une similarité distributionnelle?
- Nous examinerons la proximité distributionnelle des Nag prototypiques : les N déverbaux en *-eur/-euse/-rice*.
- Nous comparerons le comportement distributionnel des N issus des différents procédés de construction présentés dans la littérature comme formant des Nag.

1. L'agentivité nominale
2. Méthodologie
3. Identification des noms d'agent
4. Concurrence morphologique
5. Conclusion

L'agentivité nominale

Définir l'agentivité

- Définir la notion d'agent est réputé difficile (Cruse 1973, DeLancey 1984, Schlesinger 1989, Dowty 1991, Van Valin & Wilkins 1996, *i.a.*). Les conceptions varient quant à :
 - l'intentionnalité de l'agent,
 - l'animéité de l'agent,
 - la dynamicité du prédicat en jeu.
- Nous définissons *a minima* un agent comme un effectuateur (élément qui déploie l'énergie nécessaire à la réalisation d'un procès) prototypiquement animé et intentionnel.

- Deux niveaux d'instanciation de l'agentivité peuvent être distingués.
- Au niveau discursif, le rôle sémantique d'agent s'emploie pour rendre compte des relations grammaticales en contexte.
 1. [Ma sœur]_{AGENT} a assommé [le voleur]_{PATIENT}.
- Au niveau lexical, la notion d'agent peut s'employer pour caractériser le type sémantique de certains mots. Un Nag peut être défini comme un N qui décrit son référent en tant qu'agent d'une action ou d'un type d'action donné.
 2. Le nom VOLEUR est un nom d'agent en français.

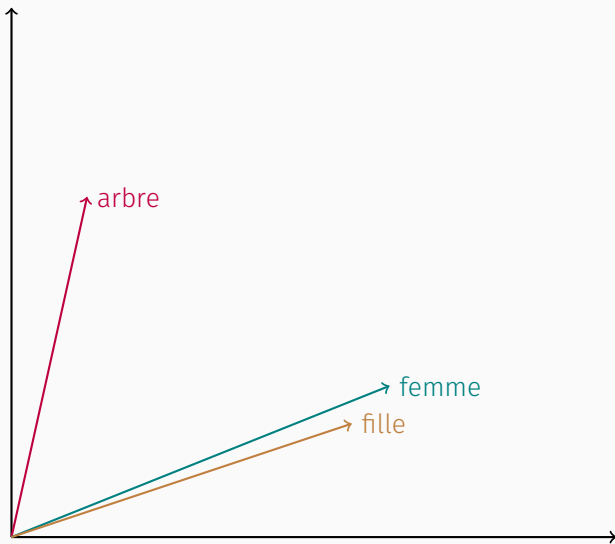
Composant actionnel

- Comment garantir la présence d'un composant actionnel dans la structure sémantique d'un N ?
- Une réponse possible est de définir comme Nag des N déverbaux exprimant l'argument agentif du V de base.
- Cette définition conduit à scinder des classes sémantiquement homogènes mais morphologiquement disparates .
 3. a. SCULPTEUR, GUÉRISSEUR, RÉDACTEUR, MANIFESTANT, PROTESTATAIRE
 - b. ARTISTE, MÉDECIN, SCRIBE, GRÉVISTE, ÉMEUTIER
- Des N non déverbaux peuvent-ils être considérés comme des Nag? Si oui, comment garantir leur agentivité, et comment délimiter l'ensemble des Nag?

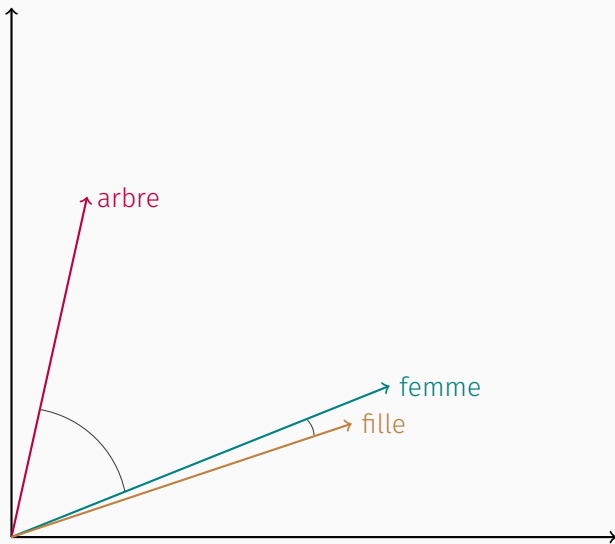
Méthodologie

- L'hypothèse distributionnelle établit une corrélation entre la similarité sémantique et le partage de contextes (Harris 1954, Firth 1957).
- Le sens des mots peut être représenté dans un espace multidimensionnel sous la forme de vecteurs calculés à partir de leurs contextes en corpus.
 - Les mots qui partagent des contextes similaires sont représentés par des vecteurs proches.
 - La nature géométrique de ces représentations permet l'application de diverses opérations : calcul de distance, addition, soustraction, etc.

Espaces vectoriels



Espaces vectoriels



Vecteurs moyens et représentation moyenne

- Comment représenter de façon unitaire les propriétés sémantiques de l'ensemble d'une classe nominale ?
- Nous construisons un vecteur unique (barycentre) à partir de la moyenne des vecteurs des N qui constituent la classe visée
(Kintsch 2001, Wauquier *et al.* 2018, Bonami & Paperno 2018)
- Nous postulons que le barycentre agrège les propriétés sémantiques définitionnelles de la catégorie.
- Les plus proches voisins du barycentre sont considérés comme des instances représentatives de la classe visée.

- Word2Vec (Mikolov et al 2013)
 - Nous utilisons l'outil dans son paramétrage par défaut (CBOW, Negative Sampling, fréquence minimum de 5, fenêtre de 5 mots)
- Corpus Wikipedia
 - Version française du 1er octobre 2018
 - 600 millions de mots
 - Parsé avec Talismane (Urieli & Tanguy 2013), lemmatisé et étiqueté

Identification des noms d'agent

Sélection de Nag prototypiques

- Nous partons de l'ensemble des N en *-eur/-euse/-rice* [N-*eur*] présents dans la base lexique.org, pour constituer un ensemble prototypique de Nag.
- 2215 N-*eur* sont recensés.
- Un tri manuel de ces N est opéré.

Sélection de Nag prototypiques

- Sont exclus de la liste retenue :
 - les N-*eur* non déverbaux (AMBASSADEUR)
 - les N-*eur* monosémiques 'instrument' (MINUTEUR)
 - les N-*eur* polysémiques 'agent/instrument' (NAVIGATEUR)
 - les N-*eur* dérivés d'un V ayant au moins une acception non dynamique (POSSESSEUR)
 - les N-*eur* dont l'un des sens agentifs au moins est fortement désémantisé par rapport à la base verbale (FACTEUR)
- Sont en revanche inclus :
 - les N-*eur* dénotant des agents non intentionnels (RONFLEUR)
 - les N-*eur* dénotant des agents non nécessairement humains (CATALYSEUR)

Construction des vecteurs

- Nous obtenons une liste de 1121 Nag. Parmi eux, 681 se rencontrent dans le corpus avec une fréquence ≥ 5 .
- Les Nag conservés sont quasiment exclusivement des N d'humain [Nhum].
- Des vecteurs sont construits pour chacun de ces Nag. Le vecteur moyen (barycentre) est également construit.
- Les plus proches voisins du barycentre dans l'espace vectoriel sont recensés.

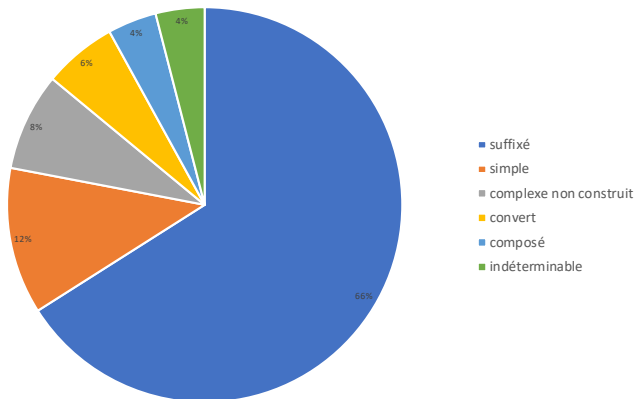
Voisinage du barycentre des Nag en -eur

Nom	Proximité		
		VOLEUR	0.692
		SOIGNEUR	0.691
		MAGICIEN	0.687
		JARDINIER	0.686
		VOYOU	0.685
		FARCEUR	0.683
		CAMBRIOLEUR	0.680
		RECELEUR	0.678
		DOMESTIQUE	0.678
		DOMPTEUR	0.677
		COLPORTEUR	0.674
		BIJOUTIER	0.670
		CRIMINEL	0.670
		POSTIER	0.669
		ASSASSIN	0.668
		...	
PLOMBIER	0.779		
TRUAND	0.751		
ESCROC	0.749		
RABATTEUR	0.737		
COIFFEUR	0.721		
PROXÉNÈTE	0.718		
PRESTIDIGITATEUR	0.714		
GANGSTER	0.704		
GARAGISTE	0.702		
PICKPOCKET	0.702		
CUISINIER	0.700		
CAMIONNEUR	0.698		
MALFRAT	0.697		
CHARLATAN	0.697		

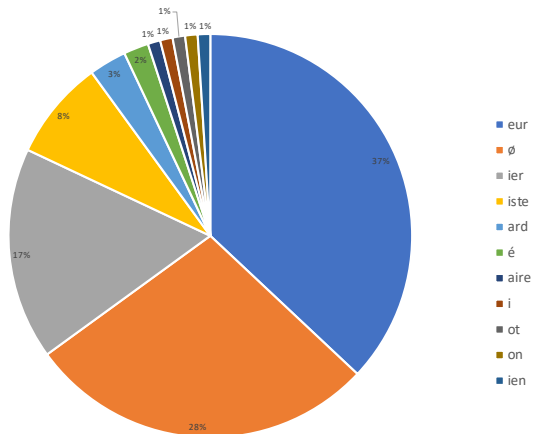
Analyse des 100 premiers voisins

- 29% des voisins ont servi à la création du barycentre.
- À une exception près (CHIEN), tous les voisins sont des Nhum.
- La plupart des voisins sont des N suffixés mais la liste comprend également des N morphologiquement simples (PROXÉNÈTE), des convertis (DROGUÉ) et des complexes non construits (IVROGNE).
- Les N construits intègrent des exposants morphologiques variés (-eur, -ier, -iste, -ard, etc.).
- Les bases de ces N sont grammaticalement variées (V, N d'action, N d'objet, Adj).

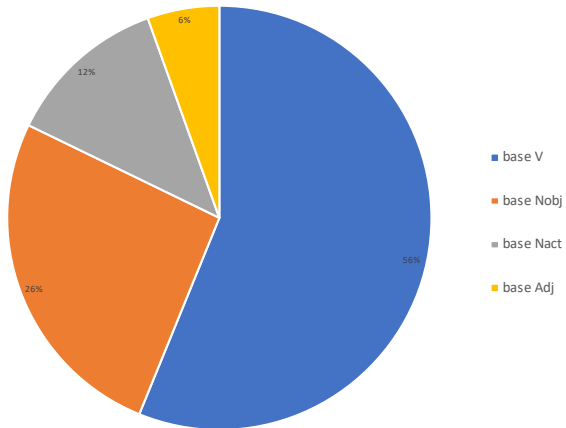
Construction morphologique des 100 premiers voisins du barycentre des Nag en *-eur*



Exposants morphologiques des 100 premiers voisins du barycentre des Nag en *-eur*



Bases des N construits parmi les 100 premiers voisins du barycentre des Nag en -eur



- Les N en *-eur* listés pour construire le barycentre comprennent deux composants sémantiques fondamentaux : (i) le trait humain et (ii) le trait agentif.
- La distribution de ces N est-elle plus fondamentalement déterminée par un trait sémantique que par un autre ?
- À titre de comparaison, nous observons les scores de proximité de Nhum non agentifs avec le barycentre des Nag en *-eur* :
 - Nhum généraux
 - Nhum phasiques
 - Nhum relationnels
 - N d'habitant (gentilés)

Proximité des Nhum généraux

Nom	Proximité
PERSONNE	0.406
HOMME	0.616
QUIDAM	0.457
GENS	0.550
GARS	0.558
GUS	0.250
GUSSE	0.298

Nom	Proximité
MONSIEUR	0.414
MECTON	0.152
MESSIEURS	0.312
SIEUR	0.263
BOUGRE	0.432
GAILLARD	0.321

Proximité des Nhum phasiques

Nom	Proximité
ADOLESCENT	0.582
ADULTE	0.424
ADO	0.432
JEUNE	0.417
TRENTENAIRE	0.442
QUINQUAGÉNAIRE	0.455
VIEILLARD	0.525

Nom	Proximité
NOURRISSON	0.382
JOUVENCEAU	0.313
FILLETTE	0.352
MIOCHE	0.250
NOUVEAU-NÉ	0.374
DONZELLE	0.423

Proximité des Nhum relationnels

Nom	Proximité
CAMARADE	0.526
PÈRE	0.494
CONJOINT	0.403
TANTE	0.283
BELLE-FILLE	0.228
OTAGE	0.359
GENDRE	0.375

Nom	Proximité
NIÈCE	0.240
FILS	0.464
COLOCATAIRE	0.527
IDOLE	0.470
COMPAGNE	0.333
AÏEUL	0.398

Proximité des N d'habitant

Nom	Proximité
CREUSOIS	0.202
NEUCHÂTELOIS	0.168
SUD-AMÉRICAIN	0.041
SLOVÈNE	0.177
RENNAIS	0.302
FIDJIEN	0.100
ARAGONAIS	0.162

Nom	Proximité
ZIMBABWÉEN	0.307
ASIATIQUE	0.381
CÉVENOLE	0.001
ACADIEN	0.264
BRUXELLOIS	0.228
TOULOUSAIN	0.342

- Les scores de proximité observés pour les Nhum non agentifs sont plus faibles que dans le cas des premiers voisins du barycentre des Nag en *-eur*.
- On peut faire l'hypothèse que la proximité avec ce barycentre est révélatrice de la combinaison des traits (i)+(ii), et donc du sens agentif des voisins.

- Afin d'évaluer l'agentivité de certains types de N, nous observons leur proximité avec les différents barycentres de Nhum :
 - N en *-eur* dénominaux
 - N en *-aire*
 - N en *-iste*
 - N en *-ien*
 - N en *-ier*
 - N en *-ant*
 - N en relation de conversion avec un V
 - N morphologiquement simples

Nom	Nag	Nphase	Nrel	Nhab
FARCEUR	0.683	0.642	0.341	0.162
AMBASSADEUR	0.361	0.111	0.416	0.160
BASKETTEUR	0.383	0.143	0.216	0.234

TABLE 1 : Proximité de N dénominaux en *-eur* avec les barycentres des Nhum

Nom	Nag	Nphase	Nrel	Nhab
STAGIAIRE	0.486	0.347	0.238	0.058
FAUSSAIRE	0.663	0.439	0.267	0.092
GESTIONNAIRE	0.399	0.118	0.072	0.035

TABLE 2 : Proximité de N en *-aire* avec les barycentres des Nhum

Nom	Nag	Nphase	Nrel	Nhab
EXORCISTE	0.569	0.453	0.361	0.062
CYCLISTE	0.405	0.261	0.137	0.326
PIANISTE	0.424	0.247	0.407	0.127

TABLE 3 : Proximité de N en *-iste* avec les barycentres des Nhum

Nom	Nag	Nphase	Nrel	Nhab
CHIRURGIEN	0.579	0.401	0.422	0.205
MÉCANICIEN	0.578	0.362	0.376	0.139
PHARMACIEN	0.572	0.420	0.422	0.275

TABLE 4 : Proximité de N en *-ien* avec les barycentres des Nhum

Nom	Nag	Nphase	Nrel	Nhab
BRANCARDIER	0.487	0.369	0.246	0.151
HORLOGER	0.541	0.365	0.363	0.299
ÉMEUTIER	0.358	0.420	0.146	0.200

TABLE 5 : Proximité de N en *-ier* avec les barycentres des Nhum

Nom	Nag	Nphase	Nrel	Nhab
ATTAQUANT	0.490	0.260	0.232	0.193
MANIFESTANT	0.283	0.383	0.040	0.106
SURVEILLANT	0.554	0.485	0.309	0.080

TABLE 6 : Proximité de candidats en *-ant* aux barycentres

Nom	Nag	Nphase	Nrel	Nhab
JUGE	0.409	0.315	0.292	0.171
ASSASSIN	0.668	0.600	0.558	0.128
GUIDE	0.369	0.214	0.129	0.102

TABLE 7 : Proximité de N convertis avec les barycentres des Nhum

Nom	Nag	Nphase	Nrel	Nhab
MÉDECIN	0.616	0.495	0.484	0.196
AVOCAT	0.522	0.341	0.478	0.216
CONCIERGE	0.487	0.479	0.507	0.083

TABLE 8 : Proximité de N morphologiquement simples avec les barycentres des Nhum

- Il existe des procédés variés de construction morphologique des Nag.
- Lorsque les Nag sont dérivés, leur base n'est pas nécessairement verbale.
- Lorsque les Nag sont dérivés de N, leur agentivité ne semble pas dépendre du fait que la base décrive une action.
 - Proximité barycentre Nag (-*eur*) déverbaux / barycentre Nag (-*eur* & -*iste*) dérivés de Nact : 0.729
 - Proximité barycentre Nag (-*eur*) déverbaux / barycentre Nag (-*eur* & -*iste*) dérivés de Nobj : 0.757
- Les Nag ne sont pas nécessairement dérivés de mots qui impliquent sémantiquement un agent.

Concurrence morphologique

- On peut s'interroger sur la coexistence de différents modes de construction des Nag en français : la variation de construction existante est-elle corrélée à une différenciation sémantique des Nag?
- Dans le but de répondre à cette question, nous comparons les comportements distributionnels des Nag en *-aire*, *-ant/-ante*, *-eur/-euse/-rice*, *-ien/-ienne*, *-ier/-ière*, *-iste*.
- Les N candidats sont sélectionnés à partir de listes extraites de la base lexique.org et contrôlées manuellement.

- Sont exclus des listes de candidats, pour chaque exposant morphologique :
 - les N non analysables comme construits en synchronie (AUTISTE)
 - les N dont l'une des acceptions au moins paraît fortement désémantisée par rapport à la base (HUISSIER)
 - les N dont la dernière étape de construction n'est pas la suffixation (TÉLÉVANGÉLISTE)
 - les N ayant au moins une acception non agentive (CANOTIER)
 - les N de type complexe dotés d'au moins une facette non agentive (IMPRESSIONNISTE)

- Sont finalement retenus (avec une fréquence ≥ 5) :
 - 28 Nag en *-aire* (DIAMANTAIRE, PLAGIAIRE, MANDATAIRE)
 - 28 Nag en *-ant* (COMMUNICANT, GOUVERNANTE, PLAIGNANT)
 - 717 Nag en *-eur* (BAIGNEUR, CHRONIQUEUSE, HOCKEYEUR)
 - 31 Nag en *-ien* (GÉNÉTICIENNE, MAGICIENNE, OPTICIEN)
 - 107 Nag en *-ier* (BOUQUETIÈRE, GUICHETIER, MEURTRIER)
 - 146 Nag en *-iste* (BASSISTE, ESSAYISTE, ORTHOPHONISTE)
- Nous analysons (i) le profil morphosémantique des Nag retenus, (ii) leur comportement distributionnel.

Analyse des N amorces : bases grammaticales

Suff.	Base V	Base N
-AIRE	6	21
-ANT	86	-
-EUR	681	36
-IEN	-	54
-IER	6	155
-ISTE	7	200

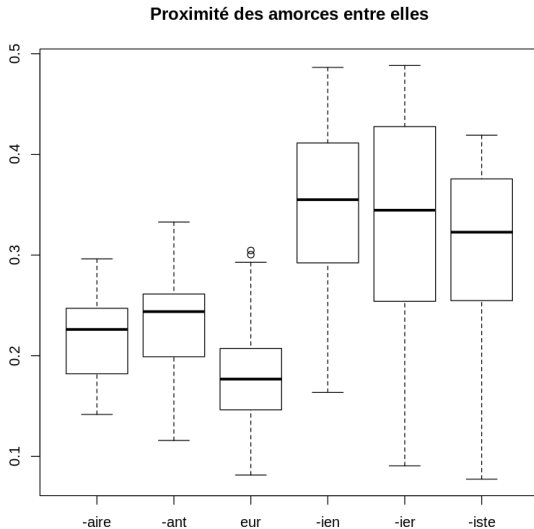
TABLE 9 : Catégorie grammaticale de la base des N amorces

Analyse des N amorces : bases sémantiques

Suff.	Action	Objet	Domaine	Obj-cog	Autre
-AIRE	19	6	-	2	-
-ANT	86	-	-	-	-
-EUR	686	12	18	1	-
-IEN	4	2	46	-	2
-IER	21	127	4	6	3
-ISTE	32	42	105	25	3

TABLE 10 : Type sémantique de la base des N amorces

Analyse des N amorces : dispersion



Voisinage des barycentres : premières observations

- Le voisinage du barycentre des *N-iste* semble constitué principalement de N qui dénotent des spécialistes de domaines et qui pratiquent des activités associées (PÉDAGOGUE, NATUROPATHE, NEUROLOGUE).
- Le voisinage du barycentre des *N-ien* est sémantiquement proche de celui des *N-iste* (exception : les N d'artiste paraissent plus proches des *N-iste*). Les N en *-logue* semblent être des voisins privilégiés des *N-iste* et des *N-ien*.

Voisinage des barycentres : premières observations

- Le voisinage du barycentre des *N-ier* paraît essentiellement constitué de N de métier manuel (FORGERON, TANNEUR, GARAGISTE). Le recouvrement entre amorces et voisins est particulièrement important.
- Le voisinage des barycentres des *N-aire* et des *N-ant* semble plus disparate que dans les cas précédents. On trouve parmi les voisins des *N-ant* un certain nombre de Nag occasionnels (CRIMINEL, TRANSFUGE, INTERVENANT).

Analyse des voisins : construction morphologique

Suff.	Suffixé	Convert	Composé	Complexe	Simple	Autre
-AIRE	57	16	6	8	12	1
-ANT	48	20	-	7	14	11
-EUR	66	7	3	8	10	6
-IEN	81	-	14	2	2	1
-IER	78	5	3	11	1	2
-ISTE	87	-	10	4	-	-

TABLE 11 : Construction morphologique des voisins des barycentres des Nag (/100)

Analyse des voisins : exposants morphologiques

Suff.	-aire	-ant	-eur	-ien	-ier	-iste	-logue	Autre
-AIRE	7	4	19	7	8	13	-	-
-ANT	1	16	12	7	6	3	-	2
-EUR	1	-	33	2	17	7	-	5
-IEN	1	-	3	17	-	28	30	-
-IER	-	2	17	1	52	5	-	2
-ISTE	1	-	4	10	-	45	26	-

TABLE 12 : Exposant morphologique des voisins des barycentres des Nag (/100)

Analyse des voisins : bases grammaticales

Suff.	Base V	Base N	Base Adj
-AIRE	28	37	8
-ANT	43	19	6
-EUR	38	30	5
-IEN	3	78	-
-IER	26	55	2
-ISTE	4	82	1

TABLE 13 : Catégorie grammaticale de la base des voisins dérivés des barycentres des Nag (/100)

Analyse des voisins : bases sémantiques

Suff.	Action	Objet	Domaine	Propriété	Obj-cog	Autre
-AIRE	33	12	12	9	4	3
-ANT	41	8	-	13	2	4
-EUR	45	16	5	5	-	2
-IEN	4	26	45	3	-	3
-IER	28	49	3	2	-	1
-ISTE	5	26	47	3	4	2

TABLE 14 : Type sémantique de la base des voisins dérivés des barycentres des Nag (/100)

Proximité moyenne des barycentres

	-AIRE	-ANT	-EUR	-IEN	-IER	-ISTE
-AIRE	-	0.762	0.780	0.732	0.744	0.715
-ANT	0.762	-	0.772	0.534	0.656	0.502
-EUR	0.780	0.772	-	0.745	0.842	0.773
-IEN	0.732	0.534	0.745	-	0.677	0.935
-IER	0.744	0.656	0.842	0.677	-	0.716
-ISTE	0.715	0.502	0.773	0.935	0.716	-

TABLE 15 : Score de proximité entre les barycentres des suffixes agentifs

- Pour les Nag en *-iste* et *-ien*, il existe une congruence entre :
 - la proximité assez élevée entre les amorces de chaque procédé de construction
 - l'homogénéité sémantique des voisins de chaque barycentre
 - la grande proximité entre les deux barycentres
- Les N-*iste* et les N-*ien* ont un comportement proche. Ils dénotent des spécialistes et des agents d'activités intellectuelles (avec une spécificité dans le domaine artistique dans le cas des N-*iste*).

- Pour les Nag en *-ier*, il existe une congruence entre :
 - la proximité entre amorces assez élevée
 - le taux de recouvrement important entre amorces et voisins
 - une nette prédilection pour les bases dénotant des objets
- Les N-*ier* se distinguent des autres Nag par le fait qu'ils dénotent très majoritairement des métiers de l'artisanat ou du commerce.

- Les Nag en *-aire*, *-ant* et *-eur* sont plus diversifiés que les précédents, comme en témoignent l'éloignement important entre les amorces de chaque suffixe et la variété morphosémantique des voisins.
 - Dans le cas de *-ant*, l'éloignement relatif du barycentre et la présence dans le voisinage de N non agentifs suggère que le suffixe ne forme pas des Nag prototypiques, ce qui est lié à une prédilection pour les Nag occasionnels.
 - Le suffixe *-aire* a la particularité de former des Nag statutaires sans préférence marquée pour un type d'occupation donnée et de sélectionner principalement comme bases des Nact.
 - Le suffixe *-eur* apparaît comme le plus productif et le plus polyvalent, comme l'indiquent la position centrale du barycentre, l'hétérogénéité des voisins et amorces, et le voisinage exclusif des Nag dispositionnels (IVROGNE, SADIQUE, ARRIVISTE).

Conclusion

Noms d'agent et hétérogénéité morphologique

- L'agentivité nominale peut être mise en évidence par l'analyse distributionnelle. La proximité avec le barycentre des Nag en *-eur* peut être considérée comme significative de l'agentivité nominale.
- Une catégorie de Nag fondée sur les propriétés distributionnelles regroupe des N morphologiquement hétérogènes. Il n'y a pas de contrainte stricte pesant sur :
 - la catégorie de la base des Nag construits,
 - le sens de la base des Nag construits,
 - le caractère construit ou non des Nag.
- Certains suffixes construisent de manière privilégiée des Nag, mais il n'y a pas de suffixe agentif unique, ni de suffixe uniquement agentif.

Diversité morphosémantique

- Certaines tendances sémantiques apparaissent pour les différents procédés de construction des Nag.
- Ces tendances corrélerent (i) l'exposant morphologique, (ii) la sélection grammaticale et sémantique des bases, (iii) le type d'agent décrit.
- Les différents procédés de construction des Nag varient quant à l'homogénéité sémantique et distributionnelle des N construits.
- La proximité sémantique entre les différents Nag varie selon les procédés de construction.

- Quelles sont les particularités sémantiques des Nag convertis, ou se terminant en *-logue,-graphe*, etc. ?
- Les propriétés de polyfonctionnalité des suffixes mobilisés dans la construction de Nag peut-elle éclairer leurs spécificités agentives ?
- La similarité distributionnelle entre les Nag dépend-elle d'autres propriétés sémantiques que celles associées à leur construction morphologique ?

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 : Mouton.
- 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 ? *Silicales* 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 : repérages des 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 2016*, 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 MacMillan, 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 déverbaux en -eur et en -ant. *Cahiers de lexicologie* 26, 35-54.

Voisinage du barycentre des N-aire

Nom	Proximité		
COMPTABLE	0.779	MAGISTRAT	0.687
MÉDECIN	0.756	JURISTE	0.685
FONCTIONNAIRE	0.746	BANQUIER	0.684
INFORMATEUR	0.721	ARMURIER	0.680
COMMERÇANT	0.718	BIBLIOTHÉCAIRE	0.677
PHARMACIEN	0.717	OUVRIER	0.671
MILITAIRE	0.713	SAVANT	0.670
EMPLOYÉ	0.712	LOBBYISTE	0.669
FINANCIER	0.710	PRATICIEN	0.665
COMMIS	0.708	AVOCAT	0.665
VÉTÉRINAIRE	0.703	INDUSTRIEL	0.663
COLLABORATEUR	0.698	ADMINISTRATEUR	0.661
OFFICIER	0.693	CITOYEN	0.660
ENTREPRENEUR	0.692	TERRASSIER	0.659
		ÉDUCATEUR	0.658
		...	

Voisinage du barycentre des N-ant

Nom	Proximité		
DIRIGEANT	0.758	DÉTENU	0.687
EMPLOYÉ	0.756	CITOYEN	0.674
RESPONSABLE	0.731	AGENT	0.672
INFORMATEUR	0.708	RETRAITÉ	0.671
COLLABORATEUR	0.705	INVESTISSEUR	0.671
LOBBYISTE	0.700	DÉCIDEUR	0.670
ÉTUDIANT	0.694	EXPERT	0.669
EMPLOYEUR	0.693	TERRORISTE	0.669
MILITANT	0.693	MILITAIRE	0.667
INTERVENANT	0.691	CLIENT	0.667
ACTIVISTE	0.690	COMPTABLE	0.665
TRAVAILLEUR	0.690	DÉLINQUANT	0.660
ANALYSTE	0.690	PRATICIEN	0.660
SYMPATHISANT	0.688	TECHNICIEN	0.655
		CRIMINEL	0.650
		...	

Voisinage du barycentre des N-ien

Nom	Proximité		
PHYSICIEN	0.840	ANTHROPOLOGUE	0.800
STATISTICIEN	0.838	OPHTALMOLOGUE	0.800
INFORMATICIEN	0.825	SOCIOLOGUE	0.794
NEUROLOGUE	0.825	CRIMINOLOGUE	0.793
LINGUISTE	0.815	ÉCONOMISTE	0.792
MATHÉMATICIEN	0.814	HÉMATOLOGUE	0.790
CHIMISTE	0.812	PSYCHOLOGUE	0.790
CARDIOLOGUE	0.808	HÉLLÉNISTE	0.782
BIOLOGISTE	0.807	ÉLECTROTECHNICIEN	0.782
PHYSIOLOGISTE	0.804	GERMANISTE	0.777
PSYCHIATRE	0.803	PÉDAGOGUE	0.775
PÉDIATRE	0.802	MÉDECIN	0.773
NEUROCHIRURGIEN	0.801	OPHTALMOLOGISTE	0.768
JURISTE	0.801	DERMATOLOGUE	0.767
		PHARMACOLOGUE	0.767
		...	

Voisinage du barycentre des N-ier

Nom	Proximité		
CORDONNIER	0.860	TANNEUR	0.783
ARMURIER	0.862	MENUISIER	0.783
BOULANGER	0.822	CHARRETIER	0.779
BIJOUTIER	0.817	GARAGISTE	0.778
MARÉCHAL-FERRANT	0.816	COIFFEUR	0.778
SERRURIER	0.812	JARDINIER	0.777
CHARPENTIER	0.808	AUBERGISTE	0.776
PLOMBIER	0.802	CAFETIER	0.774
FORGERON	0.802	QUINCAILLIER	0.772
PERRUQUIER	0.797	PORTEFAIX	0.770
ÉPICIER	0.788	CONTREMAÎTRE	0.768
BOTTIER	0.787	CUISINIER	0.767
CHAUDRONNIER	0.786	TAILLEUR	0.767
BOUCHER	0.784	MIROITIER	0.756
		VOITURIER	0.756
		...	

Voisinage du barycentre des N-iste

Nom	Proximité		
VIOLONISTE	0.800	NEUROLOGUE	0.759
CLARINETTISTE	0.798	PHONÉTICIEN	0.756
PHOTOGRAPHE	0.794	PÉDIATRE	0.755
PÉDAGOGUE	0.784	CÉRAMISTE	0.753
OPHTALMOLOGUE	0.782	ALTISTE	0.752
VIOLONCELLISTE	0.777	ORNITHOLOGUE	0.752
PIANISTE	0.775	ETHNOMUSICOLOGUE	0.752
MUSICOLOGUE	0.768	MYCOLOGUE	0.752
MUSICIEN	0.767	GERMANISTE	0.752
FLÛTISTE	0.764	LINGUISTE	0.748
ÉCRIVAIN	0.764	CLAVECINISTE	0.747
PRESTIDIGITATEUR	0.763	HAUTBOÏSTE	0.746
COMPOSITEUR	0.762	TROMPETTISTE	0.743
NATUROPATHE	0.759	NEUROCHIRURGIEN	0.742
		ILLUSTRATEUR	0.739
		...	