

Exploration de la concurrence affixale fondée sur les dictionnaires électroniques et les lexiques morphologiques

Nabil Hathout

CLLE, CNRS & Université Toulouse Jean Jaurès

Thématiques actuelles de la recherche en TAL
18 décembre 2023

En collaboration avec **Basilio Calderone**

Concurrence affixale

Identifier les affixes concurrents à partir des définitions

Identifier les affixes concurrents à partir d'embeddings lexicographiques

Conclusion

Concurrence affixale

On appelle **concurrence affixale** la situation dans laquelle un locuteur a la possibilité d'utiliser plusieurs affixes pour construire un nouveau lexème.

Si un locuteur a dans son lexique un verbe **dapotir** et qu'il veut construire un nom qui signifie '**action de dapotir**' il aura le choix de le faire au moyen d'une suffixation en :

- ▶ **-age**: **dapotissage** (*polir* → *polissage*)
- ▶ **-ment**: **dapotissement** (*ramollir* → *ramollissement*)

⇒ *-age* et *-ment* sont des affixes concurrents.

Concurrence affixale

Si son lexique contient un adjectif **grabe** et qu'il veut construire '**qualité de ce qui est grabe**', il pourra le faire par une suffixation en :

- ▶ **-ité: grabité** (*fluide* → *fluidité*)
- ▶ **-esse: grabesse** (*large* → *largesse*)
- ▶ **-erie: graberie** (*bizarre* → *bizarrerie*)
- ▶ **-itude: grabitude** (*exact* → *exactitude*)

⇒ *-ité*, *-esse*, *-erie* et *-itude* sont des affixes concurrents.

Positionnement de l'étude

La concurrence affixale est une question classique qui a fait l'objet de nombreuses études en morphologie dérivationnelle.

Nous abordons cette question d'une manière plus **expérimentale** de manière à faire émerger les affixes concurrents à partir

- ▶ des représentations sémantiques des lexèmes construits
- ▶ des annotations de ces lexèmes dans des lexiques morphologiques
- ▶ des représentations de surface (formes graphémiques) des lexèmes de base

L'étude porte sur le français, mais les méthodes et les expériences ne sont pas spécifiques au français. L'étude peut être refaite de manière strictement identique sur l'anglais et sur l'italien.

sectionIdentifier les affixes concurrents à partir de la littérature

La concurrence affixale dans la littérature

La concurrence est généralement considérée globalement au niveau du lexique complet.

Les affixes concurrents sont identifiés pour quelques classes de constructions dérivationnelles (noms d'action, noms d'agent, adjectifs de relation, etc.).

On peut facilement constituer des listes d'affixes concurrents à partir d'études morphologiques et notamment de travaux de thèses.

La concurrence affixale est forte dans les processus de transpositions catégorielles ($V \rightarrow N$ d'action, $N \rightarrow A$ de relation, $A \rightarrow N$ de qualité).

La concurrence affixale dans la littérature

La thèse de Strnadová (2014) permet de compiler plusieurs listes d'affixes concurrents qui permettent de construire des adjectifs dénominaux.

Adjectifs de relation dénominaux (N → 'relatif à N'):

-el relation → relationnel

-al artisanat → artisanal

-ique folklore → folklorique

-aire aliment → alimentaire; un régime alimentaire

-eux brume → brumeux; une visibilité brumeuse

-esque livre → livresque; une mise en page livresque

La concurrence affixale dans la littérature

Adjectifs de relation construits sur des toponymes (Strnadová, 2014) :

- ais* Islande → islandais
- ien* Canada → canadien
- ain* Afrique → africain
- ois* Lille → lillois
- in* Gironde → girondin
- an* Castille → castillan
- on* Saxe → saxon
- at* Auvergne → auvergnat

La concurrence affixale dans la littérature

La thèse de Koehl (2012) permet de compiler une liste d'affixes permettant de construire des noms de propriété désadjectivaux.

- ité* égal → égalité
- eur* ample → ampleur
- esse* étroit → étroitesse
- itude* basque → basquitude
- erie* bigot → bigoterie
- ise* ringard → ringardise

La concurrence affixale dans la littérature

L'article de Fradin (2014) énumère un ensemble d'affixes permettant de construire des noms d'action déverbaux dont :

- age** nettoyer → nettoyage
- ment** rallier → ralliement
- ion** démolir → démolition
- ade** roucouler → roucoulade
- erie** railler → raillerie
- ance** délivrer → délivrance
- ence** émerger → émergence
- ure** souder → soudure

La concurrence affixale dans la littérature

On trouve dans l'article de Huyghe & Wauquier (2021) une liste d'affixes concurrents permettant de construire des noms d'agent :

- eur** danser → danseur
- euse** chanter → chanteuse
- rice** produire → productrice
- ant** habiter → habitant
- ante** enseigner → enseignante
- aire** exécuter → exécutaire
- ier** jacasser → jacassier
- ière** jacasser → jacassière
- iste** transformer → transformiste

Objectifs de l'étude

- ▶ Comment construire des listes d'affixes concurrents à partir de ressources comme des dictionnaires, des lexiques morphologiques ou des corpus ?
- ▶ Comment identifier les affixes concurrents pour une construction particulière ?

Concurrence affixale

Identifier les affixes concurrents à partir des définitions

Identifier les affixes concurrents à partir d'embeddings
lexicographiques

Conclusion

Découvrir les affixes concurrents à partir des définitions

Lorsque deux dérivés sont sémantiquement dans le même relation avec leurs bases, mais qu'ils sont construits au moyen d'affixes différents, les deux affixes peuvent être considérés comme concurrents.

- ▶ goudron**age** = **action de** goudronner
- ▶ transform**ation** = **action de** transformer
- ▶ déplac**ement** = **action de** déplacer

Découvrir les affixes concurrents à partir des définitions

Une **définition morphologique** décrit le sens d'un mot construit relativement à un autre mot de sa famille morphologique (Martin, 1992). Le mot de la famille n'est pas toujours la base du mot construit.

- ▶ **clocheton** : petit bâtiment en forme de **clocher**, de tourelle, dont on orne les angles ou le sommet d'une construction
- ▶ **glaçon** : morceau de **glace**
- ▶ **développement** : action de **développer**, de **se développer** ou résultat de cette action, au propre et au figuré
- ▶ **productivisme** : doctrine selon laquelle la **production** est un objectif premier, système qui prône le sacrifice de toute autre considération pour maximiser la **productivité**

Les définitions morphologiques fournissent des descriptions explicites des relations sémantiques entre les lexèmes construits et leurs bases.

Découvrir les affixes concurrents à partir des définitions

Les définitions morphologiques permettent d'identifier des classes de constructions morphologiques qui correspondent à une même relation sémantique.

action de V	noms d'action déverbaux
relatif au N	adjectifs de relation dénominaux
que l'on peut V	adjectifs de capacité ou de possibilité
d'une manière A	adverbes de manière

On abstrait des **patrons de définition** à partir des définitions morphologiques dans lesquelles on remplace le ou les mots qui appartiennent à la famille dérivationnelle de l'entrée par leurs catégorie.

Les affixes qui correspondent à des définitions qui matchent un même patron sont concurrents. La concurrence est identifiée au niveau du lexique entier.

Les définitions morphologiques sont extraites du dictionnaire électronique GLAWI (Sajous & Hathout, 2015; Hathout & Sajous, 2016) qui a été créé à partir du Wiktionnaire.

- ▶ 1 966 332 entrées
- ▶ 335 487 lemmes
- ▶ 1 630 845 formes fléchies qui ne sont pas des lemmes

Nous utilisons uniquement les entrées qui sont des lemmes et qui appartiennent aux catégories majeures nom, verbe, adjectif, adverbe. Les noms propres sont exclus. Nous utilisons des définitions lemmatisées.

- ▶ 333 560 définitions
- ▶ 238 011 lemmes

Glawinette est un lexique morphologique dérivationnel construit **de manière automatique** à partir des sections morphologiques et des définitions morphologiques de GLAWI (Hathout et al., 2020). Une entrée décrit une relation entre un couple de lexèmes. Elle contient :

- ▶ le couple de lexèmes : **déplacement N / déplacer V**
- ▶ un patron d'alternance général (BAP) : **$\hat{(.+)}ment\$ / \hat{(.+)}r\$$**
- ▶ un patron d'alternance fin (FAP) : **$\hat{(.+)}ement\$ / \hat{(.+)}er\$$**
- ▶ l'origine du couple : définition ou section morphologique
- ▶ la définition morphologique (le cas échéant) : **action de déplacer ou de se déplacer .**

Glawinette contient :

- ▶ 161 116 entrées (symétriques)
- ▶ 79 167 lexèmes distincts
- ▶ 74 283 couples de lexèmes distincts
- ▶ **55 168 entrées issues de définitions morphologiques**
- ▶ 5480 couples de FAP

75% des patrons d'alternance fins sont des descriptions linguistiquement valides du procédé de construction morphologique.

Glawinette fournit les définitions morphologiques des couples issus de ces définitions. On peut extraire les FAP des lexèmes construits et les patrons de leurs définitions de cette seule ressource.

MorphyNet

MorphyNet est une collection de **lexiques** flexionnels et **dérivationnels** créés à partir des rubriques étymologiques du Wiktionnaire de 15 langues (Batsuren et al., 2021) :

allemand, anglais, catalan, espagnol, finnois, **français**, hongrois, italien, mongole, polonais, portugais, russe, serbo-croate, suédois, tchèque.

Les entrées du lexique dérivationnel fournissent :

- ▶ le lemme de la base et sa catégorie : **déplacer V**
- ▶ le lemme du dérivé et sa catégorie : **déplacement N**
- ▶ l'affixe : **ment**
- ▶ le type d'affixation : **suffix**

MorphyNet contient :

- ▶ 72 952 entrées
- ▶ 92 471 lexèmes
- ▶ 2470 affixations

Nous avons croisé les entrées de MorphyNet avec les définitions lemmatisées de GLAWI pour produire un jeu de données de lexèmes dérivés disposant d'une définition morphologique, annotés par leur lexème de base et leur affixe.

Le jeu de donnée comporte :

- ▶ 32 110 entrées
- ▶ 39 897 lemmes
- ▶ 1165 affixes

Différences entre Glawinette et MorphyNet

1. Glawinette décrit des relations directes en indirectes ; MorphyNet ne décrit que des relations directes.
2. les FAP décrivent les allomorphies radicales ; MorphyNet fournit un affixe « normalisé ».

éblouissement / éblouir

- ▶ **Glawinette** : $\text{^(.+)\text{issement}\$ / ^{(.+)\text{ir}\$}$
- ▶ **MorphyNet** : **-ment**

3. les FAP peuvent comporter un préfixe et un suffixe ; ce n'est pas le cas pour les affixes de MorphyNet.

déracinement / racine

- ▶ **Glawinette** : $\text{^dé(.+)\text{ement}\$ / ^{(.+)\text{e}\$}$
- ▶ **MorphyNet** : **dé-**

Dispositif expérimental

Nous appellerons **exposants** les FAP et les affixes.

Nous avons unifié leur notation en indiquant par **+** la position du radical :

+issement / **+ir** ; **+ment** ; **dé+ement** / **+e** ; **dé+**

Nous avons séparé les entrées en fonction des catégories grammaticales des lexèmes construits et de leurs bases. Deux procédés ne sont potentiellement concurrents que s'ils permettent de construire des dérivés de la même catégorie (cible) à partir de bases de la même catégorie (source).

Pour chaque couple de catégories, nous avons conser*vé uniquement les patrons de définitions des jeux de données Glawinette et MorphyNet

- ▶ qui apparaissent plus de 5 fois
- ▶ qui sont associés à plus d'un exposant.

Nous avons associé à chaque patron la liste des exposants qui lui correspondent avec leur fréquence.

Résultats

	patrons	nb moyen exposants
Glawinette	487	8
Morphynet	192	5

Résultats

Les listes d'exposants concurrents sont conformes aux résultats des études descriptives et à l'intuition.

N = action de V .

- ▶ **Glawinette :** +age=500 +ement=407 +ation=325 +issement=42
+e=34 +issage=26 +ification=26 +erie=23 +ion=18 +ée=15
+=15 +iement=14 +cation=14 +geage=12 +çage=11
+ition=10 +ade=10 +ellement=7 +ance=7 +èlement=6 +ure=5
+tion=5 +sion=5 +i=5 +ction=5 +èvement=4 +gation=4 ...
- ▶ **MorphyNet :** +age=536 +ment=459 +ation=316 +erie=20
+tion=17 +ion=13 +ée=11 +ance=8 +ade=8 +ition=5
+aison=4 +ure=3 +ification=3 +ie=3 +on=2 +fication=2
+faction=2 +ture=1 +ière=1 +ison=1 +eur=1 +ence=1
+ature=1 +ange=1 +aille=1

Résultats

A = relatif à le N .

- ▶ **Glawinette** : +ique=495 +iste=184 +al=55 +aire=40 +ien=35 +ier=29 +nel=26 +if=26 +tique=24 +e=24 +atoire=13 +=13 +istique=12 +el=11 +eux=10 +métrique=9 +esque=9 +ateur=9 +étique=8 +atif=8 +érique=7 +ste=7 +ntiel=7 +atique=7 +ial=6 ...
- ▶ **MorphyNet** : +ique=379 +el=27 +al=26 +aire=22 +if=15 +iste=14 +ien=6 +iel=5 +eux=4 +oire=3 +ier=3 +ial=3 +é=1 +uel=1 +istique=1 +fère=1 +en=1 +aque=1 +acé=1

N = celui qui V .

- ▶ **Glawinette** : +eur=173 +ateur=37 +sseur=17 +ant=6 +cateur=5 +seur=3 +gateur=3 +tributeur=2 +ificateur=2 +ier=2 +eux=2 +esueur=2 +aisseur=2 as+geant=1 +stituteur=1 +scripteur=1 +preneur=1 +ie=1 +cteur=1 +=1
- ▶ **MorphyNet** : +eur=135 +ateur=15 +teur=1 +ier=1

Concurrence globale

Beaucoup de patrons sont des variantes.

N = action de V .

N = action de se V .

N = action de V ou de se V .

N = action de V ; résultat de ce action .

N = action de V ou résultat de ce action .

Il est possible de réunir les ensembles d'affixes concurrents pour retrouver des concurrences plus globales.

On peut créer pour chaque couple de catégories un graphe dont les sommets sont les affixes et dont les arcs indiquent l'existence d'une concurrence entre les deux affixes.

Les arcs sont orientés et valués par la somme des fréquences des affixes pour les patrons de définitions pour lesquels ils sont concurrents.

Concurrence globale

On clusterise le graphe en utilisant par exemple **affinity propagation**.

À partir de MorphyNet

N, A, +ien

+ien +aniste

N, A, +iste

+iste +ique

N, A, +ité

+ence +ie +ise +eur +eté +esse +ité +at +isme +erie +ation +ance
+té +tude +itude +osité

Concurrence globale

À partir de MorphyNet

N, V, +age

+ence +ie +age +ade +ure +ange +ification +ison +aison +ification
+ière +ée +on +aille +ment +ation +erie +ance +isme +tion +ture
+ition +oir +ion +faction +ature

N, V, +eur

+ier +euse +teur +ant +eur +aire +ateur +ataire +ard +eux

N, V, +trice

+rice +atrice +trice

Discussion

En l'absence de jeux de données de référence, nous ne pouvons pas réaliser d'évaluation quantitative des listes que nous avons constituées.

Les listes obtenues à partir de MorphyNet sont globalement incluses dans celles obtenues à partir de Glawinette.

Les listes de Glawinette sont plus complètes (ex. conversions $V \rightarrow N$) mais plus bruitées que celles de MorphyNet (du fait des allomorphies radicales).

Concurrence affixale

Identifier les affixes concurrents à partir des définitions

**Identifier les affixes concurrents à partir d'embeddings
lexicographiques**

Conclusion

Découvrir les affixes concurrents pour une construction particulière

On peut observer la concurrence affixale pour chaque lexème construit individuellement.

L'idée est de décrire la construction à l'aide d'un ensemble de descripteurs et d'utiliser un classifieur pour prédire les affixes qui peuvent être utilisés pour réaliser cette construction.

Nous avons construit des descripteurs pour la forme du lexème de base (i.e., du lexème source) et pour le sens du lexème dérivé.

Embeddings formels

Les **descripteurs formels** sont des représentations vectorielles (*embeddings*) de la séquence des caractères du lemme de base. Les embeddings formels sont les états d'activation de l'encodeur d'un réseau de neurones LSTM bidirectionnel (seq2seq).

Nous avons utilisé la même architecture que pour Phonolette (Calderone et al., 2023).

En entrée, les caractères sont représentés par des vecteurs **one hot**.

L'encodeur contient 150 neurones.

Les représentations sont de dimension 300.

Nous avons construit deux modèles :

- ▶ un modèle entraîné sur les lemmes sources de Glawinette
- ▶ un modèle entraîné sur l'ensemble des lemmes de GLAWI

Embeddings sémantiques

Les **descripteurs sémantiques** sont des représentations de la séquence des mots qui constituent les définitions des lexèmes construits. Les embeddings sémantiques sont les états d'activation de l'encodeur d'un réseau de neurones LSTM bidirectionnel (seq2seq).

En entrée, les mots de la définition sont fournis sous forme de représentations vectorielles construites en utilisant word2vec (Mikolov et al., 2013).

Les embeddings word2vec sont créés à partir d'un corpus constitué des définitions lemmatisées des lemmes de GLAWI précédées chacune par le lemme de l'entrée de l'article.

- ▶ farine poudre blanche obtenue par la mouture de grains céréaliers et utilisée dans la fabrication du pain et de pâtisseries
- ▶ chaise siège avec dossier , sans accoudoirs
- ▶ décomposable qui peut être décomposé

Embeddings sémantiques

Hyperparamètres : taille maximale de la fenêtre = **15 mots** ; fréquence minimale des mots = **1** ; sous-échantillonnage = 0.1 ; skip-gram ; dimensions = 50

L'encodeur contient 150 neurones.

Les représentations sont de dimension 300.

Nous avons construit deux modèles :

- ▶ un modèle entraîné sur les définitions morphologiques de Glawinette
- ▶ un modèle entraîné sur l'ensemble des définitions de GLAWI

Embeddings sémantiques

Les jeux de données sont réduits aux entrées dont les définitions contiennent **entre 3 et 15 mots**.

jeu de données	entrées
Glawinette	50 793
MorphyNet	21 020

Le jeu de données produit à partir de MorphyNet contient uniquement les entrées

- ▶ qui disposent d'une représentation vectorielle de leur forme de surface et
- ▶ qui disposent d'une définition morphologique

Dispositif expérimental

La concurrence existe uniquement entre les affixes qui construisent des dérivés de la même catégorie cible à partir de la même catégorie source.

- ▶ Le jeu de données est divisé en 16 sous-ensembles, un pour chaque couple de catégories.
- ▶ Les sous-jeux de données dont les lexèmes dérivés sont construits par moins de 5 affixes sont exclus.
- ▶ On ne considère que les 30 affixes les plus fréquents dans les entrées du sous-jeu (ou moins le cas échéant).

Pour chaque couple de catégories, nous avons entraîné un classifieur **XGBoost** (valeurs des paramètres par défaut) sur la totalité du sous-jeu de données en utilisant les descripteurs formels et sémantiques ou les descripteurs sémantiques seulement.

Les modèles sont utilisés pour prédire les 5 affixes les plus probables pour chaque entrée du sous-jeu de données correspondant.

Dispositif expérimental

La probabilité de l'affixe (i.e., d'appartenance à la classe) est enregistrée.

La précision des modèles entraînés sur la forme et le sens sur les jeux d'entraînement est très élevée.

Glawinette (F+S)

cible	A	N	R	V
A	0.98	0.99	0.99	0.98
N	0.96	0.97	0.99	0.95
R	1.00			
V	0.98	0.94	1.00	0.92

MorphyNet (F+S)

cible	A	N	R	V
A	0.97	0.97		0.99
N	0.92	0.97		0.97
R				
V	0.99	0.81		0.93

Glawinette (S)

cible	A	N	R	V
A	0.98	0.99	0.98	0.98
N	0.90	0.93	0.99	0.90
R	1.00			
V	0.98	0.88	1.00	0.83

MorphyNet (S)

cible	A	N	R	V
A	0.97	0.97		0.99
N	0.92	0.97		0.97
R				
V	0.99	0.81		0.92

Résultats Glawinette

broderie: action de broder .

F+S +erie=0.476 +age=0.437 +ement=0.058 +ure=0.008
+issement=0.007

S +erie=0.357 +age=0.295 +ement=0.278 +ation=0.0235
+issage=0.020

pillage: action de piller , ou dégât qui en être le suite .

F+S +age=0.621 +ement=0.123 +ation=0.120 +erie=0.056
+e=0.033

S +age=0.800 +ement=0.149 +ation=0.029
+issement=0.007 +issage=0.003

Résultats Glawinette

éplucheur: celui , celle qui éplucher .

F+S +eur=0.987 +ateur=0.004 +ant=0.003 +sseur=0.001
+age=0.000

S +eur=0.985 +ateur=0.007 +sseur=0.003 +ant=0.001
+=0.000

utilisatrice: celle qui utiliser un outil ou quelque_chose .

F+S +atrice=0.951 +eur=0.017 +ation=0.006 +ateur=0.006
+ement=0.005

S +atrice=0.929 +eur=0.035 +ement=0.010 +oir=0.003
+erie=0.003

Résultats MorphyNet

broderie: action de broder .

F+S +age=0.603 +ment=0.207 +erie=0.131 +ation=0.051
+ure=0.001

S +age=0.475 +erie=0.248 +ment=0.221 +ation=0.046
+ure=0.002

pillage: action de piller , ou dégât qui en être le suite .

F+S +ment=0.428 +ation=0.348 +age=0.180 +erie=0.029
+ance=0.003

S +ment=0.431 +ation=0.321 +age=0.190 +erie=0.041
+tion=0.003

Résultats MorphyNet

éplucheur: celui , celle qui éplucher .

F+S +eur=0.965 +ateur=0.027 +ard=0.002 +ant=0.001
+eux=0.000

S +eur=0.967 +ateur=0.026 +ard=0.002 +ant=0.001
+eux=0.000

utilisatrice: celle qui utiliser un outil ou quelque_chose .

F+S +euse=0.778 +eur=0.2042 +trice=0.003 +ade=0.002
+age=0.001

S +euse=0.631 +eur=0.344 +trice=0.004 +ateur=0.004
+age=0.002

Résultats MorphyNet

Les affixes concurrents correspondent aux prédictions dont la probabilité est suffisamment élevée (≥ 0.100)

Nombre de prédictions dont la probabilité est supérieure à 0.100

Glawinette (F+S)

nb	AA	AN	AR	AV	NA	NN	NR	NV	RN	VA	VN	VR	VV
1	511	1664	926	1620	3133	3905	246	2785	6	2329	2319	18	1971
2	59	283	667	346	3313	4085	7	2835	0	704	2661	1	4767
3	26	121	83	30	1531	2920	2	430	0	124	1727	0	408
4	6	38	2	4	451	1056	0	29	0	16	279	0	67
5	2	3	0	1	52	174	0	1	3	0	18	0	8

MorphyNet (F+S)

nb	AA	AN	AV	NA	NN	NV	VA	VN	VV
1	364	989	493	1246	2280	1535	1225	1439	1266
2	65	209	32	767	1265	480	34	1541	1116
3	26	29	0	357	577	17	0	1598	128
4	8	5	0	73	133	0	0	105	3

Discussion

Les concurrents sont déterminés individuellement pour chaque dérivé. Une comparaison avec les réponses de sujets pour les mêmes lexèmes construits devient ainsi possible.

La méthode permet d'estimer la « disponibilité » des concurrents. Cette disponibilité ne dégrade pas la sélection des affixes. Des propriétés sémantiques (plus fines que les définitions) permettent de discriminer entre les affixes concurrents.

Les propriétés de surface ont peu d'effet sur les prédictions. Les prédictions sont fondées principalement sur le sens.

Les embeddings de forme utilisés pour construire le jeu de données MorphyNet ne sont probablement pas suffisamment précis.

Concurrence affixale

Identifier les affixes concurrents à partir des définitions

Identifier les affixes concurrents à partir d'embeddings lexicographiques

Conclusion

Le dispositif des expériences qui utilisent des embeddings peut être amélioré en utilisant des embeddings formels qui a rendent mieux compte des similarités morphologiques. L'objectif sera d'augmenter l'informativité des descripteurs formels.

Nous pourrions concevoir une tâche de *wugs* dérivationnels dans laquelle des locuteurs produiraient des dérivés de pseudo-mots qui permettraient de tester la concurrence entre un ensemble d'affixes (ex., entre des affixes d'adjectif relationnels déverbaux).

Nous envisageons de refaire les mêmes expériences sur l'anglais et sur l'italien. Il deviendrait alors possible de comparer le phénomène de concurrence affixale dans différentes langues.

Nous prévoyons d'adapter le dispositif

- ▶ en utilisant des embeddings statiques comme description du sens des lexèmes construits (i.e., remplacer les embeddings de définitions par des embeddings des lexèmes dérivés) pour comparer les descriptions des acceptions que sont les définitions avec les descriptions distributionnelles ;
- ▶ en utilisant des embeddings contextuels comme représentation sémantique des définitions morphologiques.

Conclusion

Les dictionnaires électroniques sont des ressources riches en informations morphologiques qui peuvent être utilisés pour explorer la question de la concurrence affixale.

La concurrence peut être observée au niveau du lexique (i.e., d'une classe de dérivés comme les noms d'agent), d'une relation morphosémantique particulière ou d'un lexème construit particulier.

À l'époque des *foundation models* (modèles de fondation), les embeddings d'objets particuliers (définitions de dictionnaires) restent pertinents pour des études spécifiques.

Cette étude confirme les résultats de Wauquier (2020); Wauquier et al. (2020, 2021) qui ont mis en évidence que

Les exposants sont fortement discriminés par le sens des dérivés.

Références

- Batsuren, Khuyagbaatar, Gábor Bella & Fausto Giunchiglia. 2021. MorphyNet: a large multilingual database of derivational and inflectional morphology. In *Proceedings of the 18th sigmorphon workshop on computational research in phonetics, phonology, and morphology*, 39–48.
- Calderone, Basilio, Nabil Hathout & Olivier Bonami. 2023. Phonolette: a grapheme-to-phoneme converter for French. *Lexique* 57–70.
- Fradin, Bernard. 2014. La variante et le double. In Sophie David, Sarah Leroy & Florence Villoing (eds.), *Foisonnements morphologiques. etudes en hommage à Françoise Kerleroux*, 111–148. Nanterre: Presses Universitaires de Paris Ouest.
- Hathout, Nabil & Franck Sajous. 2016. Wiktionnaire's Wikicode GLAWified: a workable French machine-readable dictionary. In *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*, Portorož, Slovenia.

Références

- Hathout, Nabil, Franck Sajous, Basilio Calderone & Fiammetta Namer. 2020. Glawinette: a linguistically motivated derivational description of French acquired from GLAWI. In *Proceedings of the twelfth international conference on language resources and evaluation (LREC 2020)*, 3870–3878. Marseille.
- Huyghe, Richard & Marine Wauquier. 2021. Distributional semantics insights on agentive suffix rivalry in French. *Word Structure* 14(3). 354–391.
- Koehl, Aurore. 2012. *La construction morphologique des noms désadjectivaux suffixés en français*. Nancy: Université de Lorraine Thèse de doctorat.
- Martin, Robert. 1992. *Pour une logique du sens* Linguistique nouvelle. Paris: Presses universitaires de France.
- Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. 1–12.
- Sajous, Franck & Nabil Hathout. 2015. GLAWI, a free XML-encoded Machine-Readable Dictionary built from the French Wiktionary. In *Proceedings of the of the eLex 2015 conference*, 405–426. Herstmonceux, England.

- Strnadová, Jana. 2014. *Les réseaux adjectivaux : sur la grammaire des adjectifs dénominaux en français*. Université Paris Diderot / Univerzita Karlova V Praze Thèse de doctorat.
- Wauquier, Marine. 2020. *Confrontation des procédés dérivationnels et des catégories sémantiques dans les modèles distributionnels*. Université Toulouse Jean Jaurès Thèse de doctorat.
- Wauquier, Marine, Cécile Fabre & Nabil Hathout. 2020. Semantic discrimination of technicality in French nominalizations. *Zeitschrift für Wortbildung / Journal of Word Formation* 2020(2). 100–121.
- Wauquier, Marine, Cécile Fabre & Nabil Hathout. 2021. Différenciation des noms d'action dérivés : le facteur de technicité étudié en corpus. In Cécile Frérot & Mojca Pecman (eds.), *Des corpus numériques à l'analyse linguistique en langues de spécialité*, 241–261. UGA éditions.