

Exploration computationnelle de la concurrence affixale

Nabil Hathout, Basilio Calderone

CLLE, CNRS & Université Toulouse Jean Jaurès

Thématiques actuelles de la recherche en TAL
14 octobre 2024

Concurrence affixale

Identifier les affixes concurrents à partir de la littérature

Identifier les affixes concurrents à partir des définitions

Exploitation des erreurs de prédiction par des modèles neuronaux

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 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* → *bizarrierie*)
- -itude : **grabitude** (*exact* → *exactitude*)

→ -ité, -esse, -erie et -itude sont concurrents

Concurrence affixale

La concurrence affixale est plus forte entre les précédés qui servent à la **transposition** :

V → N d'action -age, -ment, -ion, -ure, -erie, -ance

N → A de relation -ique, -el, -aire, -eux, -ier, -ien

A → N de qualité -ité, -eur, -itude, -erie, -esse

Il n'y a pas de concurrence pour les adverbes de manière

Un seul affixe est disponible : -ment

Positionnement de l'étude

La concurrence affixale fait l'objet de nombreuses études en morphologie dérivationnelle

Pour un panorama récent sur la question, voir Huyghe & Varvara (2023)

Nous abordons cette question d'une manière **computationnelle** pour faire émerger les affixes concurrents à partir

- ▶ des annotations des lexèmes construits dans un lexique dérivationnel
- ▶ de représentations sémantiques et formelles de ces lexèmes et de leurs bases

L'étude porte sur le français

Les méthodes et les expériences présentées ne sont pas spécifiques au français

Concurrence affixale

Identifier les affixes concurrents à partir de la littérature

Identifier les affixes concurrents à partir des définitions

Exploitation des erreurs de prédiction par des modèles neuronaux

Conclusion

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 dans la littérature

La thèse de Strnadová (2014) permet de compiler plusieurs listes d'affixes concurrents permettant 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

Concurrence affixale

Identifier les affixes concurrents à partir de la littérature

Identifier les affixes concurrents à partir des définitions

Exploitation des erreurs de prédiction par des modèles neuronaux

Conclusion

Découvrir les affixes concurrents à partir des définitions

Lorsque deux dérivés sont sémantiquement dans la 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 complexe relativement à un autre mot de sa famille morphologique (Martin, 1992)

- ▶ **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**

Le mot de la famille qui motive le défini n'est pas toujours sa base

- ▶ **cléricaiser** : inspirer l'esprit de **cléricanisme** → rendre **clérical**

Découvrir les affixes concurrents à partir des définitions

Les définitions morphologiques fournissent des descriptions explicites des relations sémantiques entre les glexèmes complexes et les lexèmes qui les motivent

Elles permettent d'identifier des classes de procédés 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

Découvrir les affixes concurrents à partir des définitions

On abstrait des **patrons de définitions** à partir des définitions morphologiques dans lesquelles on remplace le mot qui appartient à la famille dérivationnelle de l'entrée par sa catégorie

glissabilité	: capacité de	glisser	sur un support
N	: capacité de	V	sur un support

Les affixes qui correspondent à des définitions quiinstancient le même patron sont concurrents

La concurrence est identifiée au niveau du lexique entier

Dictionnaire électronique GLAWI

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 des catégories majeures nom, verbe, adjectif, adverbe

Les noms propres sont exclus

Les définitions sont lemmatisées

- ▶ 238 011 lemmes
- ▶ 333 560 définitions

Glawinette (fra_GLN_1.4)

Glawinette est un lexique morphologique dérivationnel construit **de manière automatique** à partir des définitions et des sections morphologiques de GLAWI (Hathout et al., 2020)

Une entrée décrit une relation entre un couple de lexèmes

- ▶ les lemmes des lexèmes et leurs catégories :

déplacement N / déplacer V

- ▶ leurs exposants dérivationnels :

^(.+)ement\$ / ^(.+)er\$

- ▶ la définition morphologique (le cas échéant) :

action de déplacer ou de se déplacer .

Glawinette (fra_GLN_1.4)

Nous utilisons uniquement les entrées de Glawinette munies de définitions morphologiques :

- ▶ 53 547 couples de lexèmes munis d'une définition
- ▶ 53 242 couples de lexèmes distincts
- ▶ 66 289 lexèmes distincts
- ▶ 3 107 couples d'exposants distincts

Glawinette est à notre connaissance la seule ressource dérivationnelle du français qui fournit des **descriptions complètes** :

- ▶ formelles (lemme)
- ▶ catégorielles (cat)
- ▶ sémantiques (définition)
- ▶ morphologiques (FAP)

Autres caractéristiques de Glawinette

Glawinette décrit des relations directes (**diviniser** / **divin**) et indirectes (**détacheur** / **tache**: tache → tacher → détacher → détacheur)

Les FAP peuvent être des exposants dérivationnels ou flexionnels
lavable / **laver**: $\hat{(.+)}able\$$ / $\hat{(.+)}er\$$

Les FAP peuvent contenir des allomorphies radicales
éblouissement / **éblouir**: $\hat{(.+)}issement\$$ / $\hat{(.+)}ir\$$

Les FAP peuvent simultanément comporter un préfixe et un suffixe
déracinement / **racine**: $\hat{dé(.+)}ement\$$ / $\hat{(.+)}e\$$

La notation des FAP peut être simplifiée en indiquant par **+** la position du radical

+able / **+er**; **+issement** / **+ir**; **dé+ement** / **+e**

Dispositif expérimental (à partir des définitions)

Le jeu de données est composé des entrées dont

- ▶ les patrons de définitions apparaissent au moins 5 fois
- ▶ les patrons de définitions sont associés à 2 exposants ou plus

Les entrées du jeu de données sont séparées relativement aux catégories grammaticales des couples de lexèmes (Guzmán Naranjo & Bonami, 2023)

- ▶ Il n'y a pas de concurrence entre des précédés dont les catégories sources et cibles sont différentes

Le jeu de données contient 411 patrons.

En moyenne, un patron être réalisé par 8 exposants (la concurrence affixale est fréquente)

Résultats

Les concurrents sont spécifiques à chaque patron de définition

Les concurrents sont conformes aux descriptions (littérature) et à l'intuition

V = V à_nouveau .

re+=2732 ré+er=874 ré+r=69 r+r=28 r+er=25 ress+er=4 ress+r=3
ra+r=1

N = action de V .

+age=476 +ement=385 +ation=290 +issement=37 +e=27
+ification=26 +issage=25 +erie=23 +ion=14 +iement=14
+geage=12 +cation=12 +=10 +ée=9 +çage=9 +ade=8 +ition=7
+ance=7 +ellement=6 +èlement=5 +ction=5 +èvement=4 +ure=4
+sion=4 +gation=4 +cage=4 +aison=4 ...

Résultats

R = de une manière A .

+ment=501 +ement=112 +eusement=94 +ivement=64
+ellement=41 +amment=20 +mment=15 +ément=5 +ièrement=5
+ûment=4 +ètement=4 +ttement=3 ...

A = qui V .

+ant=244 +eur=114 +ateur=92 +issant=32 +atif=15 +isseur=12
+geant=10 +eux=8 +çant=7 +ificateur=7 +isant=5 +if=5 +ent=5
+atoire=5 +utif=4 +ucteur=4 +cateur=4 +ard=3 ...

A = relatif à la N .

+ique=249 +nel=17 +if=12 +e=12 +aire=11 +tique=10 +atoire=10
+ier=8 +al=8 +el=7 +étique=6 +érique=5 +ien=5 +atif=5
+ntiel=4 +icien=4 +ateur=4 +eux=3 +ant=3 ...

Concurrence globale

Beaucoup de patrons de définitions sont des variantes

Les variantes ont les mêmes concurrents

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 .

Les variantes sont en partie induites par les allomorphies radicales

+ement +issement +iement +ellement +èlement +èvement
+èchement +ètement +ttement

Concurrence globale

Les concurrents des différents patrons peuvent être clusterisés pour obtenir des concurrences plus globales

Pour chaque couple de catégories, on crée un graphe dont

- ▶ les sommets sont les affixes
- ▶ les arcs indiquent l'existence d'une concurrence entre les deux affixes

Les arcs sont orientés et valués

Si e est l'arc $a1 \rightarrow a2$

$$val(e) = \sum_{p \in P(a1) \cap P(a2)} nb(p, a2)$$

où $P(\alpha)$ est l'ensemble des patrons de définitions que réalise l'affixe α , et où $nb(\pi, \alpha)$ est le nombre d'entrées dans lesquelles α réalise le patron de définition π

Concurrence globale

Le graphe est clusterisé en utilisant **affinity propagation**

On obtient 67 clusters

→ Ce nombre est du même ordre de grandeur que les combinaisons de traits morphosyntaxiques en français

A, A, ultra+

ultra+ +issime hyper+ extra+ super+ archi+ multi+ sur+

A, V, in+able

in+able im+able irré+able irre+able in+issable in+çable in+ible
irr+able in+geable

R, A, +ment

+ment +ement +eusement +ivement +ellement +amment +mment
+ièrement +ément +ûment +ttement +ètement +ensément +nnement
+èrement

Concurrence globale

N, V, +age

+age +ement +ation +issement +e +ification +issage +erie +ion
+iement +cation + +geage +çage +ée +ade +ance +ellement
+ition +ure +èlement +sion +ction +aïson +cage +gation +èvement
+tion +i +èchement +pression +ètement +gage +ence +ersion
+ention +mission +te +ature +scription +issance +élation +bat
+ession +ttement +issure +rection +enue +ture +aille +ing +ication
+ène +ie +antise +is +tien +çure auto+ation +é +isme +clusion
+riture +ment +ue

Discussion

La clusterisation permet de réunir les concurrents qui réalisent des patrons similaires

Les exemplaires des clusters sont globalement pertinents

Les clusters permettent d'identifier une grande partie des allomorphies radicales

Une révision manuelle des clusters est envisageable

La révision permettrait d'associer à chaque exposant FAP un représentant

Vers une normalisation des définitions

Nous réalisons le même clustering sur les patrons de définitions

Pour chaque couple de catégories, on crée un graphe dont

- ▶ les sommets sont les patrons de définitions
- ▶ les arcs indiquent l'existence d'affixes qui réalisent deux patrons différents

Les arcs sont orientés et valués

Si e est l'arc $p1 \rightarrow p2$

$$val(e) = \sum_{a \in A(p1) \cap A(p2)} nb(p2, a)$$

où $A(\pi)$ est l'ensemble des affixes que réalisent le patron π

Vers une normalisation des définitions

Le graphe est clusterisé en utilisant **affinity propagation**

On obtient 50 clusters. → Ce nombre est du même ordre de grandeur que les combinaisons de traits morphosyntaxiques en français

N = celui qui V .

celui qui V .	personne qui V .	celui , celle qui V .
quelqu'un qui V .	ouvrier qui V .	celui ou celle qui V .
machine à V .	machine qui V .	appareil qui V .
machine pour V .	élément qui V .	qui V .
instrument pour V .		

R = de une manière A .

de une manière A .	de manière A .	de façon A .
de une façon A .	en A lieu .	de un point de vue A .

V = V à_nouveau .

V à_nouveau .	V de_nouveau .	se V à_nouveau .
V .	V une nouveau fois .	V une second fois .
variante de V .	se V de_nouveau .	

Vers une normalisation des définitions

N = action de V .

action de V .

action de V ou résultat de cette action .

action de V , résultat de cette action .

action de V ; résultat de cette action .

action de se V .

État de ce qui être V .

faire de V .

action de V ou état de ce qui être V .

action de V

action de V ou de se V .

action de V , de se V .

l' action de V .

action de V , de se V ou résultat de cette action .

action de V ; état de ce qui être V .

État de celui qui être V .

faire de être V .

ce qui V .

État de une personne V .

personne (de sexe féminin) V .

action de V , ou résultat de cette action .

ce qui servir à V .

action de (se) V .

faire de se V .

Vers une normalisation des définitions

N = action de V .

action de se V , résultat de cette action .

V .

action de V ; ou résultat de cette action .

action de V ou de se V ; résultat de cette action .

État de ce qui V .

le fait de V .

ce qui être V .

femme V .

outil qui servir à V .

personne V .

Les clusters peuvent être utilisés pour normaliser une partie des définitions au moyen de l'exemple

action de se V , résultat de cette action . → action de V .

Discussion

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

La méthode est globalement effective. Elle permet de constituer des listes d'affixes concurrents

Paradigmes dérivationnels

La clusterisation des définitions rend visible l'existence d'un **nombre réduit des relations sémantiques** qui structurent le lexique dérivationnel

- ▶ Cette observation renforce la plausibilité d'une structuration paradigmatique de la morphologie dérivationnelle

Concurrence affixale

Identifier les affixes concurrents à partir de la littérature

Identifier les affixes concurrents à partir des définitions

Exploitation des erreurs de prédiction par des modèles neuronaux

Conclusion

Prédiction des affixes

La concurrence affixale peut être identifiée en utilisant des modèles neuronaux.

L'idée est de considérer les prédictions erronées des modèles comme des indices d'hésitation entre plusieurs affixes.

Deux possibilités a priori : faire prédire aux modèles des lemmes complets (**déplacement**) ou seulement des affixes (**+ement**).

Prédiction des affixes

Les erreurs dans la prédiction des lemmes complets peut se situer au niveau du radical, de l'afixe ou des deux

- ▶ **bninquebalant** / brinquebalant
- ▶ **discrttion** / discrétisation
- ▶ **quttional** / quotientiel

Les erreurs dans la prédiction des affixes sont directement exploitables. Une partie des prédictions erronées ne sont pas des affixes.

- ▶ +éen / +iforme
- ▶ +ant / +atoire
- ▶ +ellement / +ellerie
- ▶ +erie / **+eete**

Prédiction des affixes

Nous avons construit des modèles LSTM (long-short term memory) **sequence2sequence** capables de prédire l'exposant (FAP2) d'un lexème cible construit à partir d'un lexème source

La prédiction se fait relativement aux propriétés du lexème source

- ▶ sa catégorie
- ▶ ses propriétés formelles : lemme, radical, exposant (FAP1)

Le lexème cible est identifié par

- ▶ sa catégorie
- ▶ la relation sémantique qui le relie au lexème source

- ▶ **déplacer** → **+ement**
V, (déplacer; déplac; +er)
N, 'action de V'
- ▶ **laver** → **+able**
V, (laver; lav; +er)
A, 'que l'on peut V'
- ▶ **tache** → **dé+eur**
N, (tache; tach; +e)
N, 'qui élimine le N'

Prédiction des affixes

Nous avons construit un modèle pour chacune des trois propriétés formelles de la source : M_{lem} , M_{rad} , M_{fap}

Les 3 modèles ont exactement la même architecture

Les modèles se composent de **3 encodeurs LSTM** qui encodent séparément la description formelle de la source, les catégories source et cible et la relation sémantique entre la source et la cible

Les représentations produites par les encodeurs sont concaténées et fournies à **1 décodeur LSTM** qui prédit le FAP2 la cible

Nous utilisons une méthode **10-fold cross-validation**

- ▶ le jeu de données est divisé en 10 parties
- ▶ la prédiction des entrées d'1 des parties est obtenue en utilisant un modèle entraîné sur les 9 autres parties

Prédiction des affixes

Le modèle qui a la meilleure exactitude est celui dont l'entrée formelle est FAP1

modèle	exactitude
M_{lem}	0.476
M_{rad}	0.448
M_{fap}	0.503

Les prédictions erronées des modèles qui sont des FAP2 sont potentiellement des concurrents affixaux.

Pour chaque couple de catégories

- ▶ on collecte pour chaque FAP2 attendu l'ensemble des prédictions erronées
- ▶ on ne conserve que les prédictions qui existent comme FAP2 pour le couple de catégories

Prédiction des affixes (à partir de lem)

V, N, +age

+ement=360 +ation=87 +e=38 +erie=37 +ure=12 +iement=9
+oir=8 +ade=8 +=8 +èlement=6 +é=5 +ée=4 +ttement=3
+ion=3 +euse=3 +eur=3 +cation=3 +ance=3 +ément=2
+èvement=2 +ètement=2 +te=2 +aison=2

V, N, +ation

+age=84 +ement=57 +e=10 +ance=7 +ion=5 +ateur=4 +erie=3
+abilité=3 +ée=2 +ure=2 +sion=2 +pression=2 +ition=2 +ence=2
+ative=2 +ant=2 +a=2

→ Les concurrents semblent spécifiques à chaque affixe : certains facteurs, notamment la famille, semblent déterminer la sélection des affixes

Prédiction des affixes (à partir de lem)

V, A, +able

+ant=17 +atif=5 **in**+able=4 +**c**able=4 +é=3 +u=3 +**ss**able=3
+ible=3 +**g**able=3 +eux=3 +eur=2 +e=2

→ Une partie des concurrents sont en réalité des variantes qui incluent une allomorphie radicale

→ La correspondance entre sens et forme est souple

- ▶ +able n'a normalement pas de concurrents
- ▶ +able semble exceptionnellement utilisé pour construire des adjectifs agentifs (**secourable**, **admirable**, **irritable**)

N, A, +ique

+iste=56 +al=34 +eux=31 +ien=29 +e=26 +tique=24 +é=23
+ier=21 +aire=21 +iforme=9 +istique=8 +ifère=7 +oïde=6 +in=6
+nier=5 +if=5 +esque=5 +ctique=5 +aque=5 +ial=4 +el=4
+atique=4 +ant=4 +=4 +isé=3 +ïque=2 +étique=2 +ène=2 +t=2
+stique=2 +ste=2 +sien=2 +ois=2 +iel=2 +icole=2 +icide=2

Prédiction des affixes (à partir de rad)

Les prédictions à partir du radical sont similaires aux prédictions à partir des lemmes

V, N, +age

+ement=330 +ation=55 +e=32 +erie=30 +issement=16 +ée=11
+ure=11 +=9 +ance=6 +ification=5 +ade=5 +çage=4 +oir=4
+ant=4 +issage=3 +euse=3 +eur=3 +é=2 +ture=2 +te=2
+ition=2 +ellement=2 +cation=2 +cage=2 +aison=2

V, N, +ation

+age=68 +ement=62 +ion=7 +ance=7 +e=6 +ateur=5 +ition=4
+ification=4 +abilité=4 +ée=3 +issement=3 +erie=3 +at=3 +=3
ré+ation=2 +ure=2 +sion=2 +isation=2 +et=2 +ature=2 +ade=2

Prédiction des affixes (à partir de rad)

V, A, +able

+ant=12 in+able=7 +eur=6 +é=5 +geable=5 +atif=5 +=3 +eux=2
+e=2 +çable=1 +u=1 +isable=1 +ificateur=1 +if=1 +ier=1
+ible=1 +gable=1 +ent=1 +cable=1 +ateur=1 +ard=1 +al=1

N, A, +ique

+iste=86 +al=39 +ien=31 +é=30 +e=29 +aire=21 +eux=19
+ier=16 +if=9 +iforme=8 +atique=7 +=7 +istique=6 +icole=6
+atoire=6 +atif=6 +in=5 +ifère=5 +icien=5 +étique=4 +tique=4
+ophile=4 +ologique=4 +nel=4 +ial=4 +esque=4 +érien=3 +isé=3
+isant=3 +ieux=3 +en=3 +érique=2 +énique=2 +éen=2 +t=2
+oïde=2 +orial=2 +ois=2 +nier=2 +itique=2 +iel=2 +eur=2
+aque=2

Prédiction des affixes (à partir de fap)

V, N, +age

+ement=214 +ation=150 +e=11 +erie=10 +eur=7 +é=5 +oir=5
+ée=3 +ure=3 +ant=3 +ence=2 +ance=2 +aison=2

→ Réduction importante du nombre des variantes allomorphiques

Amélioration très importante de la « précision »

V, N, +ation

+age=390 +ement=185 +e=16 +ure=14 +oir=6 +erie=4 +ant=4
+é=2 +ition=2 +eur=2 +ette=2

V, A, +able

+ant=18 +eur=3 +é=2 +atoire=2 +atif=2 in+able=1 im+able=1
+issant=1

Prédiction des affixes (à partir de fap)

N, A, +ique

+eux=64 +iste=61 +ien=56 +al=48 +é=38 +aire=28 +e=26
+ier=20 +iel=12 +nel=8 +ifère=7 +iforme=6 +atique=6 +esque=5
+=5 +ois=4 +isant=4 +yeux=3 +tique=3 +if=3 +ieux=3 +ial=3
+el=3 +aque=3 +oïde=2 +ologue=2 +istique=2 +iflore=2 +i=2
+en=2 +ant=2

L'utilisation des erreurs de prédiction des modèles qui disposent du FAP1 en entrée améliore la qualité des listes de concurrents candidats

À faire : Les concurrents peuvent être regroupés en clusters en utilisant *affinity propagation*

Concurrence affixale

Identifier les affixes concurrents à partir de la littérature

Identifier les affixes concurrents à partir des définitions

Exploitation des erreurs de prédiction par des modèles neuronaux

Conclusion

Conclusion

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

La ressource GLN fournit une caractérisation sémantique des relations dérivationnelles, en plus de la caractérisation formelle, catégorielle et morphologique

GLN renouvelle l'exploration de différents phénomènes morphologiques comme la dérivation ou la concurrence affixale

Conclusion

La concurrence affixale peut être identifiée par différentes méthodes :

- ▶ directement à partir des définitions morphologiques
- ▶ en prédisant les concurrents au moyen d'un **classifieur xgboost** à partir des lemmes sources et d'embeddings des définitions des lexèmes cibles, comme présenté dans (Hathout, 2023)
- ▶ en exploitant les erreurs de modèles neuronaux LSTM

Les concurrences mises en évidence par les différentes méthodes sont convergentes

Une combinaison des résultats des différentes méthodes pourrait permettre d'obtenir des listes de concurrents plus « propres »

Une révision manuelle des listes obtenues pourrait permettre de constituer une **description de référence de la concurrence affixale**

Conclusion

La concurrence pourrait permettre d'avancer dans l'exploration des paradigmes dérivationnels

- ▶ Les **clusters de définitions** pourraient servir à identifier les dimensions sémantiques qui structurent ces paradigmes
- ▶ Les **clusters d'affixes** permettent d'identifier une partie des allomorphies radicales (et des radicaux qui identifient les familles paradigmatiques)
- ▶ La combinaison des clusters de définitions et d'affixes pourraient servir à identifier des classes dérivationnelles

Références

- Fradin, Bernard. 2014. La variante et le double. In Sophie David, Sarah Leroy & Florence Villoing (eds.), *Foisonnements morphologiques. études en hommage à Françoise Kerleroux*, 111–148. Nanterre: Presses Universitaires de Paris Ouest.
- Guzmán Naranjo, Matías & Olivier Bonami. 2023. A distributional assessment of rivalry in word formation. *Word Structure* 16(1). 87–114.
- Hathout, Nabil. 2023. Exploration de la concurrence affixale fondée sur les dictionnaires électroniques et les lexiques morphologiques. Séminaire « Thématiques actuelles de la recherche en TAL ». CLLE. Toulouse.
- 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.
- 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.

Références

- Huyghe, Richard & Rossella Varvara. 2023. Affix rivalry: Theoretical and methodological challenges. *Word Structure* 16(1). 1–23.
- 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.
- 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.