

# Etudier les familles morphologiques à partir des lexèmes de basse fréquence. L'exemple de la dérivation verbale en français

Fabio Montermini & Juliette Thuilier

Université Toulouse Jean Jaurès & CLLE-ERSS

Sémantique pour les ressources en morphologie dérivationnelle  
8 juillet 2019



# Introduction

Compétition entre procédés morphologiques dans la formation des verbes à partir de noms et d'adjectifs en français (Roger, 2003; Tribout, 2010; Lignon, 2013; Bonami and Thuilier, 2018)

	Base nominale	Base adjectivale
<i>-iser</i>	CAMEL ~ CAMELISER	CONCRET ~ CONCRÉTISER
<i>-ifier</i>	MUSÉE ~ MUSÉIFIER	CLAIR ~ CLARIFIER
<i>en-</i>	PILE ~ EMPILER	RICHE ~ ENRICHIR
<i>a-</i>	LIGNE ~ ALIGNER	PLAT ~ APLATIR
conversion	COURBE ~ COURBER	MÛR ~ MÛRIR

- Exclusion des procédés impliquant une lecture négative ou réversative (*é-*, *dé-*)
- Le projet : déterminer quels facteurs formels et sémantiques conditionnent le choix d'un procédé à partir des verbes construits de fréquence 1
- Aujourd'hui : exploration qualitative des données de fréquence 1 en relation avec la question des familles morphologiques

# Plan

- 1 Introduction
- 2 Nos données
- 3 Que trouve-t-on dans les basses fréquences ?
- 4 Familles & paradigmes morphologiques
- 5 Conclusion

## Quelles données ?

- Lexèmes de très basse fréquence
  - ▶ Accès à des lexemes créés ‘sur le vif’, moins influencés par la lexicalisation, les changements sémantiques, la perte de transparence, la polysémie. . . ;
  - ▶ Taille des données raisonnable pour une annotation manuelle;
  - ▶ Annotation syntaxico-sémantique d’occurrences réelles, plutôt que de lexèmes abstraits.

## Nos données : extraction

- Verbes construits ayant une fréquence 1 dans FrWac (Baroni et al. 2009)
- Extraction automatique des formes lemmatisées *verbe* présentant les affixes *-iser*, *-ifier*, *en-*, *a-*
- Extraction automatique des verbes sur la base d'un lexique des N et A (GLÀFF, Sajous et al. 2013)
- Nettoyage manuel (86% des données éliminées)
- 655 items en tout.

conversion	<i>en-</i>	<i>a-</i>	<i>-ifier</i>	<i>-iser</i>
205	28	0	48	374
(31.3%)	(4.3%)	(0%)	(7.3%)	(57.1%)

# Nos données : premières observations

- Propriétés phonologiques (Lignon, 2013; Bonami and Thuilier, 2018)
  - ▶ concurrence conditionnée par la longueur du radical :
    - radicaux d'une syllabe favorisent *-ifier* et *en-*;
    - radicaux de 2 syll. ou plus favorisent *-iser*
    - conversion n'est pas particulièrement sensible à la longueur du radical
  - ▶ choix de l'affixe conditionné par le phonème initial ou final :
    - obstruante alvéolaire finale favorise *-ifier*;
    - sonante finale favorise *-iser*
    - initiale vocalique défavorise *en-*

# Nos données : premières observations

- Propriétés morphologiques :

- ▶ la structure de la famille morphologique affecte le choix du procédé (Bonami and Thuilier, 2018)
  - base adjectivale uniquement → *-ifier*
  - bases adjectivale et nominale → *-iser*
  - base nominale uniquement → conversion

- Propriétés sémantiques :

- ▶ la télicité du verbe construit conditionne le choix du procédé
  - env. 65% des verbes atéliques construits par conversion
  - env. 70% des verbes téliques construits par suffixation

- 1 Introduction
- 2 Nos données
- 3 Que trouve-t-on dans les basses fréquences ?**
- 4 Familles & paradigmes morphologiques
- 5 Conclusion



# Que trouve-t-on dans les basses fréquences ?

## 1 des termes techniques

(1) AJAXIFIER, MAILLOCHER, CARBOXYLASER /  
CARBOXYLER

## 2 un V qui comble un trou dans une famille morphologique, dont plusieurs occasionalismes

(2) CHANTAGER, MAGIFIER, CANAPISER, PÉDANTIFIER /  
PÉDANTER / PÉDANTISER

(3) ELVISPRESLER, JAMESBROWNISER,  
LAGARDEMICHARDISER

## 3 des doublons : lexèmes attestés dans frWac, ayant une base commune (surabondance)

- ▶ 27.5% dans nos données

## Remplissage d'une case dans une famille

- Spécialisation sémantique :

(4) CENTRIFIER :

*Que je sache, ce n'est pas vers le PS (que certains accusent pourtant de se "**centrifier**") que se dirigent les transfuges de l'UDF."*

## Remplissage d'une case dans une famille

- Spécialisation sémantique : atélique > télique

(5) RYTHMISER :

*Nous n'avons qu'une nature animale au départ et pourtant il nous est possible de la transformer, de la **rythmiser**, c'est-à-dire d'accorder les rythmes de notre organisme aux aspirations de notre esprit."*

## Remplissage d'une case dans une famille

- Spécialisation sémantique : télique / atélique

- (6) TALIBANER :  
*Alors les Taliban, on les laissait **talibaner** dans leur coin, lapider, torturer..."*
- (7) TALIBANISER :  
*Il est en train de reveiller les Zaouiates pour **talibaniser** l'Algérie"*
- (8) CHOCHOTTER :  
*Très amusant, donc, quand on y pense... d'avoir l'air de **chochotter**..."*
- (9) CHOCHOTTISER :  
*la mythique série de Capcom qui aurait tendance à se **chochottiser** avec un gameplay plus ouvert parait-il."*

# Surabondance

- Deux cas de figure principaux

- 1** Plusieurs verbes de basse fréquence en concurrence pour remplir une case (env. 25%)

WIKIPÉDIFIER, WIKIPÉDIER, WIKIPÉDER  
ONOMATOPER, ONOMATOPÉISER

- 2** Un lexème est créé pour une case déjà remplie (env. 75%)

SANIFIER ~ ASSAINIR

PACTER ~ PACTISER

ENCHEVILLER ~ CHEVILLER

CARICATURISER ~ CARICATURER

# Surabondance

- Dans le type 1, surtout surabondance phonologique (variation inter-locuteurs, Roché and Plénat (2014)) :
  - ▶ COUCOUER vs. COUCOULER, COUCOUTER ;
  - ▶ CRADIFIER vs. CRADER, CRADOÏFIER, CRADOSSER... ;
  - ▶ CROQUIER vs. CROQUITER ;
  - ▶ PARANOTER vs. PARANOÏER, PARANOÏDER

# Surabondance

- Dans le type 2,
- Réalignement de famille :
  - ▶ AUTOMATIQUER, CONSULTATER, VIOLENCER;
- Réalignement de série :
  - ▶ FACILISER, POURRIFIER, RIGIFIER;
- Réalignement de famille + de série :
  - ▶ COMMENTARISER, MATERNALISER, PRIORITARISER;
  - ▶ EXHIBITIONNER, INSCRIPTIONNER, SUPERPOSITIONNER;
  - ▶ ROTATIONNER, TRANSACTIONNER;

- 1 Introduction
- 2 Nos données
- 3 Que trouve-t-on dans les basses fréquences ?
- 4 Familles & paradigmes morphologiques**
- 5 Conclusion



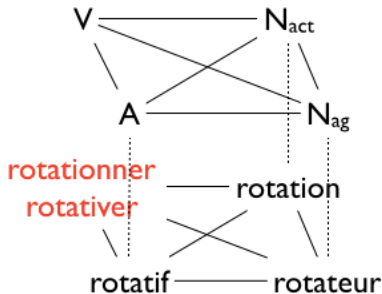
# Familles / paradigmes morphologiques

- Quelle articulation entre les données que nous observons et les familles / paradigmes morphologiques?
- Bonami and Strnadová (2018) :
  - ▶ une famille morphologique est constituée de tout ensemble de lexèmes reliés morphologiquement (forme/contenu) de manière systématique
  - ▶ un paradigme morphologique est constitué d'un ensemble de familles (superposables) qui comportent les mêmes relations de contenu.
- Parmi les forces à l'œuvre dans la compétence morphologique (contraintes), il y a la tendance à avoir des familles / paradigmes maximalement remplis et les plus cohérents possibles.

## Famille / Paradigme V-NAction-...

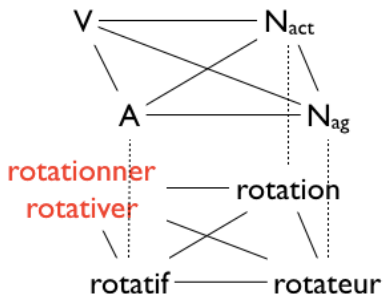
- Les cas comme ROTATIONNER constituent un premier exemple d'homogénéisation formelle au sein d'une famille:

(10) "c'est bien tout ça mais avec la mémoire que j'ai, j'ai jamais réussi à retenir les raccourcis des softs (à part ceux de bases comme déplacer, **rotationner** etc...)"



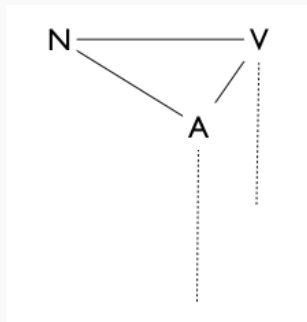
## Famille / Paradigme V-NAction-...

- (11) "Il le fait **rotativer** comme un jongleur faisant tourner ses disques du bout de ses baguettes."

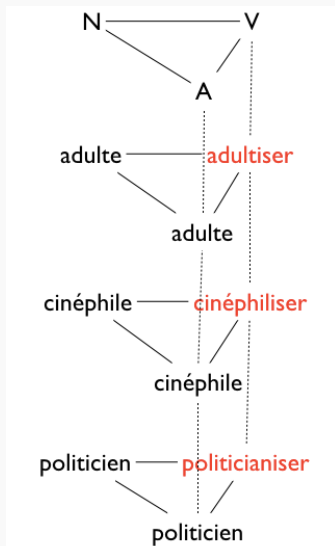
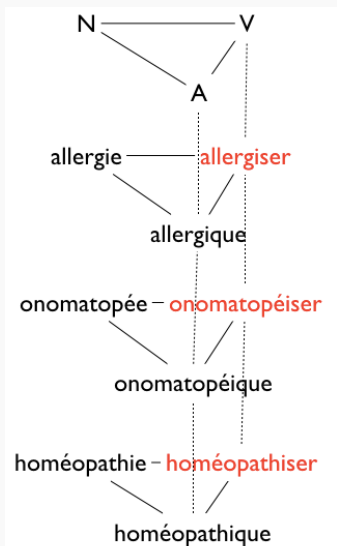


## Famille / paradigme 'objet'

- Les cas de remplissage les plus typiques concernent les familles structurées autour d'un 'objet' (Roché, 2017; Fradin, in press);
- La base correspond à une activité (FREESTYLER), à un événement (CHANTAGER), à un instrument (PISTOLETER), à un agent (MATRIARCHER)...



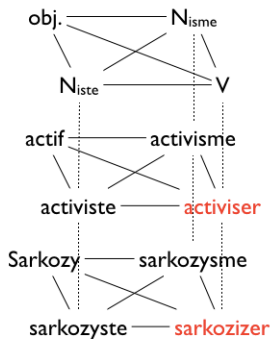
## Famille / paradigme 'objet'



## Familles axiologiques

- Certains verbes dénominaux remplissent également une case dans une famille de type axiologique (Roché, 2017) (syncrétisme);

(12) "*organiser les avant-gardes communistes combattantes dans le Parti, **activer** les mouvements de masse révolutionnaires, organiser les avant-gardes de lutte dans les organisations de masse révolutionnaires.*"



## V dénominaux causatifs

- Les cas les plus fréquents dans notre base de données sont ceux (moins typiques) de V causatifs (transformation ou mouvement) ; la base désigne (le plus souvent) le point final du processus :
  - ▶ Qualité : CRADIFIER, FÉBRILISER, RUDIFIER ;
  - ▶ Objet résultant : CHENALISER, EMBUISSONNER, LIASSER ;
  - ▶ Objet résultant (métaphorique) : AYRAULTISER, PASSOIRISER [du Mesrine], SHERIFISER ;
  - ▶ Ambiguïtés : CARTEPOSTALISER.

## Conclusions et perspectives

- Le lexique de basse fréquence (en comparaison avec le lexique 'établi') peut donner accès à la compétence morphologique active des locuteurs ;
- Il pose des questionnements méthodologiques et théoriques :
  - ▶ statut des occasionalismes ;
  - ▶ statut des doublons (avec lexèmes fréquents et rares) ;



## Conclusions et perspectives

- Proposer une analyse globale de la dérivation verbale dénominale et désadjectivale du français :
  - ▶ 4 procédés employant des stratégies différentes (préfixation, suffixation, conversion) ;
  - ▶ interaction des facteurs formels, syntaxiques et sémantiques ;
  - ▶ modélisation statistique de la compétition.
- Concordance de comportement avec le lexique 'établi' du point de vue de la forme (Bonami and Thuilier, 2018) (cf. journée du 7 juin) ;
- Annotation sémantique fines des lexèmes de basse fréquence en emploi.

## Conclusions et perspectives

- La dynamique des familles / paradigmes morphologiques peut contribuer à l'interprétation / annotation sémantique des dérivés de basse fréquence
- ex.: les ambiguïtés que nous avons repérées peuvent dériver de l'appartenance simultanée d'un lexème à plusieurs familles (synchrétisme) :
  - ▶ ex. BISTOURISER, MARTEAUPIQUERISER : famille 'activité' ou 'causatif' ;
  - ▶ ex. ACTIVISER, SARKOSIFIER : famille 'causatif' ou 'axiologique' ;
  - ▶ ACTIVISER est une forme surabondante dans la famille de ACTIF-ACTIVITÉ, mais remplit une case vide dans la famille de ACTIVISME-ACTIVISTE.

**Merci!**

## References I

- Baroni, M., Barnardini, S., Ferraresi, A., and Zanchetta, E. (2009). 'The wacky wide web : a collection of very large linguistically processed web-crawled corpora'. *Language Resources and Evaluation*, 43:209226.
- Bonami, O. and Strnadová, J. (2018). 'Paradigm structure and predictability in derivational morphology'. *Morphology*, 29:167–197.
- Bonami, O. and Thuilier, J. (2018). 'A statistical approach to rivalry in lexeme formation: French *-iser* and *-ifier*'. *Word Structure*, 11.
- Fradin, B. (in press). 'On the variety of derivational paradigms'. In *Word Formation Theories III*.
- Lignon, S. (2013). '*-iser* and *-ifier* suffixation in French: Verifying data to 'verize hypotheses''. In N. Hathout, F. Montermini, and J. Tseng (eds.), *Morphology in Toulouse. Selected Proceedings of Décembrettes 7 (Toulouse, 2-3 December 2010)*. Munich: Lincom Europa, 119–132.

## References II

- Roché, M. and Plénat, M. (2014). 'Le jeu des contraintes dans la sélection du thème présuffixal'. In F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefer, and S. Prévost (eds.), *Actes du 4e Congrès Mondial de Linguistique Française*. Paris: Institut de Linguistique Française, 1863–1878.
- Roché, M. (2017). 'Les familles dérivationnelles : comment ça marche ?'
- Roger, C. (2003). 'Derived change-of-state verbs in french: a case of semantic equivalence between prefixes and suffixes'. *Acta Linguistica Hungarica*, 1-2:187–199.
- Sajous, F., Hathout, N., and Calderone, B. (2013). 'GLÀFF, un Gros Lexique À tout Faire du Français'. In *Actes de la 20e conférence sur le Traitement Automatique des Langues Naturelles*. Les Sables d'Olonne, 285298.

## References III

Tribout, D. (2010). *La conversion de nom à verbe et de verbe à nom en français*. Ph.D. thesis, Université Paris Diderot (Paris 7).