

# From Démonette to Phononette:

## A derivational database for research in French morphophonology

Fiammetta Namer<sup>1</sup>, Stéphanie Lignon<sup>2</sup>, Nabil Hathout<sup>3</sup>

<sup>1,2</sup>UMR ATILF, Université de Lorraine,

<sup>3</sup>UMR CLLE, Université de Toulouse Jean Jaurès

BCL, Prague  
September 19, 2024

# Introduction

- We use a derivational database for French to analyze morphophonological variations in word families
  - epenthesis, apophony, consonant change, suppletion, etc.
  - we compare the phonological transcription of stems.
- What can a database with large coverage tell us about:
  - the most frequent alternation patterns?
  - new criteria contributing to rank them according to transparency and/or predictability?

# Démonette2.0 (<https://demonette.fr/>)

- Derivational database, paradigm-based framework of derivation
- Complementary content of two coindexed tables: table of relations (TR) and table of lexemes (TL)

$W_1$	$W_2$	Cat <sub>1</sub>	Cat <sub>2</sub>	WF <sub>1</sub>	Der. Patt <sub>1</sub>	WF <sub>2</sub>	Der. Patt <sub>2</sub>	Complexity	Direction
<i>menteur</i> 'lier <sub>m</sub> '	<i>menteuse</i> 'lier <sub>f</sub> '	Nm	Nf	suf	Zeur	suf	Zeuse	simple	indirect
<i>menteur</i>	<i>mentir</i> 'lie'	Nm	V	suf	Zeur		Z	simple	des2as
<i>menteur</i> 'lying'	<i>mentir</i>	Adj	V	suf	Zeur		Z	simple	des2as

Relations in TR form word families (222,118 wordpairs)

# Table of Lexemes (~ 380.000 entries)

lemma	Inflectional Paradigm	Infl. Paradigm: phonemic transcription	stem space
<i>menteur</i> <sub>A</sub>	Afpms:menteur; Afpmp:menteurs; Afpfs:menteuse; Afpfp:menteuses	Afpms:mã.tœʋ; Afpmp:mã.tœʋ; Afpfs:mã.tøz; Afpfp:mã.tøz	
<i>menteur</i> <sub>N</sub>	Ncms:menteur; Ncmp:menteurs	Ncms:mã.tœʋ; Ncmp:mã.tœʋ	
<i>mentir</i> <sub>V</sub>	Vmip1s-:mens; Vmcp1s-: mentirais; Vmcp2s-:mentirais; Vmcp3s-:mentirait; Vmif1s-: mentirai; Vmif2s-:mentiras; Vmif3s-:mentira; Vmis3p-: mentirent; Vmif2p-:mentirez; Vmcp2p-:mentiriez; [...]	Vmip1s-:mã; Vmcp1s-:mã.ti.ʋε; Vmcp2s-:mã.ti.ʋε; Vmcp3s-: mã.ti.ʋε; Vmif1s-:mã.ti.ʋε; Vmif2s-:mã.ti.ʋa; Vmif3s-: mã.ti.ʋa; Vmis3p-:mã.tiʋ; Vmif2p-:mã.ti.ʋε; Vmcp2p-: mã.ti.ʋje; [...]	mãt; māt; mã; māt; mã; māt; māt; māt; mãti; mãti; mãti; mãti

# Table of Lexemes: inflectional paradigms

lemma	Inflectional Paradigm	Infl. Paradigm: phonemic transcription	stem space
<i>menteur</i> <sub>A</sub>	Afpms:menteur; Afpmp:menteurs; Afpfs:menteuse; Afpfp:menteuses	Afpms:mã.tœʁ; Afpmp:mã.tœʁ; Afpfs:mã.tøz; Afpfp:mã.tøz	
<i>menteur</i> <sub>N</sub>	Ncms:menteur; Ncmp:menteurs	Ncms:mã.tœʁ; Ncmp:mã.tœʁ	
<i>mentir</i> <sub>V</sub>	Vmip1s-:mens; Vmcp1s-: :mentirais; Vmcp2s-:mentirais; Vmcp3s-:mentirait; Vmif1s-: :mentirai; Vmif2s-:mentiras; Vmif3s-:mentiraient; Vmip2s-: :mentiriez; Vmip3s-:mentiraient; :mentirent; Vmif2p-:mentiriez; Vmcp2p-:mentiriez	Vmip1s-:mã; Vmcp1s-:mã.ti.ʁe; Vmcp2s-:mã.ti.ʁe; Vmcp3s-: :mã.ti.ʁe; Vmif1s-:mã.ti.ʁe; Vmif2s-:mã.ti.ʁa; Vmif3s-: :mã.ti.ʁe; Vmif3p-:mã.ti.ʁe; Vmip2p-:mã.ti.ʁe; Vmcp2p-: :mã.ti.ʁe; [...]	mãt; māt; mã; māt; mã; māt; māt; māt; mãti; mãti; mãti; mãti

Source: GLÀFF lexicon of wordforms compiled from the French Wiktionary

# Table of Lexemes: stem spaces

lemma	Inflectional Paradigm	Infl. Paradigm: phonemic transcription	stem space
<i>menteur</i> <sub>A</sub>	Afpms:menteur; Afpmp:menteurs; Afpfs:menteuse; Afpfp:menteuses	Afpms:mã.tœʁ; Afpmp:mã.tœʁ; Afpfs:mã.tøz; Afpfp:mã.tøz	Structured set of the stems occurring in paradigms
<i>menteur</i> <sub>N</sub>	Ncms:menteur; Ncmp:menteurs	Ncms:mã.tœʁ; Ncmp:mã.tœʁ	
<i>mentir</i> <sub>V</sub>	Vmip1s-:mens; Vmcp1s-: mentirais; Vmcp2s-:mentirais; Vmcp3s-:mentirait; Vmif1s-: mentirai; Vmif2s-:mentiras; Vmif3s-:mentira; Vmis3p-: mentirent; Vmif2p-:mentirez; Vmcp2p-:mentiriez; [...]	Vmip1s-:mã; Vmcp1s-:mã.ti.ʁe; Vmcp2s-:mã.ti.ʁe; Vmcp3s-: mã.ti.ʁe; Vmif1s-:mã.ti.ʁe; Vmif2s-:mã.ti.ʁa; Vmif3s-: mã.ti.ʁa; Vmis3p-:mã.ti.ʁ; Vmif2p-:mã.ti.ʁe; Vmcp2p-: mã.ti.ʁje; [...]	mãt; māt; mã; māt; mã; māt; māt; māt; mãti; māti; mãti; māti

# Phononette0.1

- Phononette = TL and TR are merged;
- String distance is computed between the word stems of each derivational entry.
  - It is from the stem pair with the smallest distance that stem variation is examined
- Only Phononette entries with variations are kept: 19,032 derivational relations.
- Manual validation: 34.3% of the variations have been checked and revised, i.e., 6,540 entries.
  - Phononette0.1 is only a testbench!!!
- General distribution of alternations
- Classification attempt of alternation patterns

<b>adjunction (cf. also deletion)</b>	vowel	33.16%	<i>stable/stabil-ité</i> [stabl] / [stabil]	'stable' / 'stability' Xbl/Xbil
	consonant	11.04%	<i>bazar/bazard-er</i> [bazar] / [bazard]	'mess' / 'throw away' X/Xd
	rhyme	10.06%	<i>vociférer/vociférat-eur</i> [vosifer] / [vosiferat]	'vituperate' / 'vituperater' X/Xat
	syllable	0.16%	<i>cor/cornet-ier</i> [kɔr] / [kɔrnɛt]	'horn' / 'hornist' X/Xnɛt
<b>Feature change</b>	apophony	1.25%	<i>accidentel/accidental-ité.</i> [aksidãtɛ] / [aksidãta]	'accident' / 'accidental' XɛC/XaC
	denazalisation	29.48%	<i>bonbon/bonbonn-ière</i> [bõbõ] / [bõbɔn]	'candy' / 'candybox' Xõ/Xɔn
	assibilation & other cons. changes	1.60%	<i>accéder/access-ible</i> [aksɛd] / [aksɛs]	'access' / 'accessible' Xd/Xs
<b>Not a phonological rule</b>	partial stem suppletion	1.84%	<i>erreur/erron-é</i> [erœr] / [erɔn]	'error'/'wrong' Xœr/Xɔn
	complete stem suppletion	11.41%	<i>cheval/hipp-ique</i> [ʃɛval] / [ip]	'horse' / 'equestrian' X/Y



# Classifying stem alternation

- Similarity between inflectional and alternating stems, reflecting their distance from transparency and thus straightforward recognition by speakers
- Traditionally, variation patterns are ranked in a continuum based on string distance measures
- From "strong" stem suppletion...

- Complete and phonologically unpredictable stem difference

• <i>eau</i> / <i>év-ier</i>	vs	<i>cendre</i> /	<i>cendr-ier</i>
• 'water' / 'sink'		'ash' /	'ash tray'
• [o] / [ev-je]		[sãdr] /	[sãdr-ije]

Carstairs, 1988  
Boyé 2006, Dressler 2015

# Measuring and classifying stem alternation

- ... to regular morpho-phonologically conditioned allomorphy
  - Difference affecting usually only one phoneme (feature)

• <i>clair</i> / <i>clar-té</i>	vs	<i>pur</i> / <i>pur-eté</i>
• 'clear' / 'clarity'		'pure' / 'purity'
• [klɛr] / [klar-te]		[pyr] / [pyr-te]
• <i>bois</i> / <i>bois-erie</i>	vs	<i>pierre</i> / <i>pierr-erie</i>
• 'wood' / 'wood panelling'		'stone' / 'gem'
• [bwa] / [bwaZ-əri]		[pjɛr] / [pjɛr-əri]

Kisparsky, 1996, Tranel, 1981

# String similarity is not sufficient

- *eau* / *év-ier* vs *cheval* / *hipp-ique*
- [o] / [ev-je]                      [ʃəval] / [ip-ik]

- *amitié* / *in-imitié* vs *clair* / *clar-té*  
'friendship' / 'hostility'
- [a<sup>h</sup>mitje] / [in-i<sup>h</sup>mitje]              [klɛr] / [klar-te]

- ⇒ [ʃəval]/[ip] seems "less" suppletive than [o]/[ev]
- ⇒ Speakers associate [ip] to *cheval*

- ⇒ Phoneme insertion / alternation is predictable at various degrees

- Combine with other criteria
- Rely on Phononette

Corbin, 1985, 1987  
Plénat, 2008

# Stem integrity

C1 String similarity

C2 Stem integrity preservation

- Alternation at stem-exponent boundary preferred

- Sometimes, C2 contradicts intuition
- C1 and C2 apply at word level

- [arbr] / [arbor-e]  
(*arbre* 'tree' / *arbor-é* 'tree filled')

- [pɛl] / [pɛlt-e]

- [stabl] / [stabili-ite]  
(*stable* / *stabil-ité*)

- [klu] / [kluv-jɛr]  
(*clou* 'nail' / *clouv-ière* 'nail box')

# Criteria based on type frequency

C3 Pattern frequency (related to morphophonological series)

- How many different relations with the same alternation pattern?

C4 Stem propagation in the family

- How are variations distributed in word families

• Xbl / Xbil > X / Xv

• [jəval] / [ip]: *hippique*,  
*hippisme*, *hippodrome*,  
*hippopotame*

• [o] / [ev]: *évier*

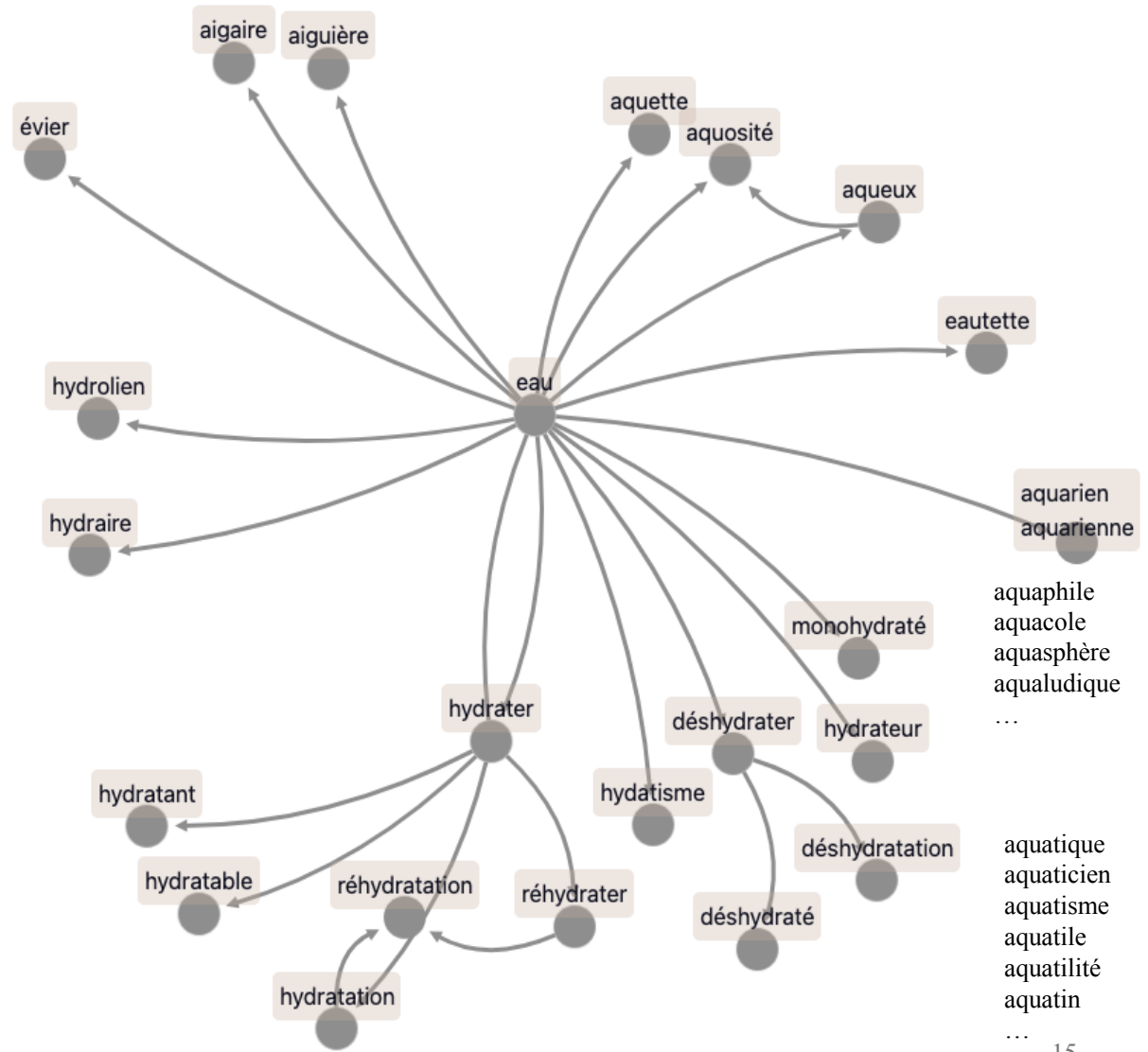
Frei, 1929  
Mel'cuk, 2006

# Impact of each criterion?

- How do criteria interact?
- What is the impact of morpho-phonological patterns? of word families?
- (Are derivational paradigms important?)
- We examine three types of stem variation, each with equal string similarity

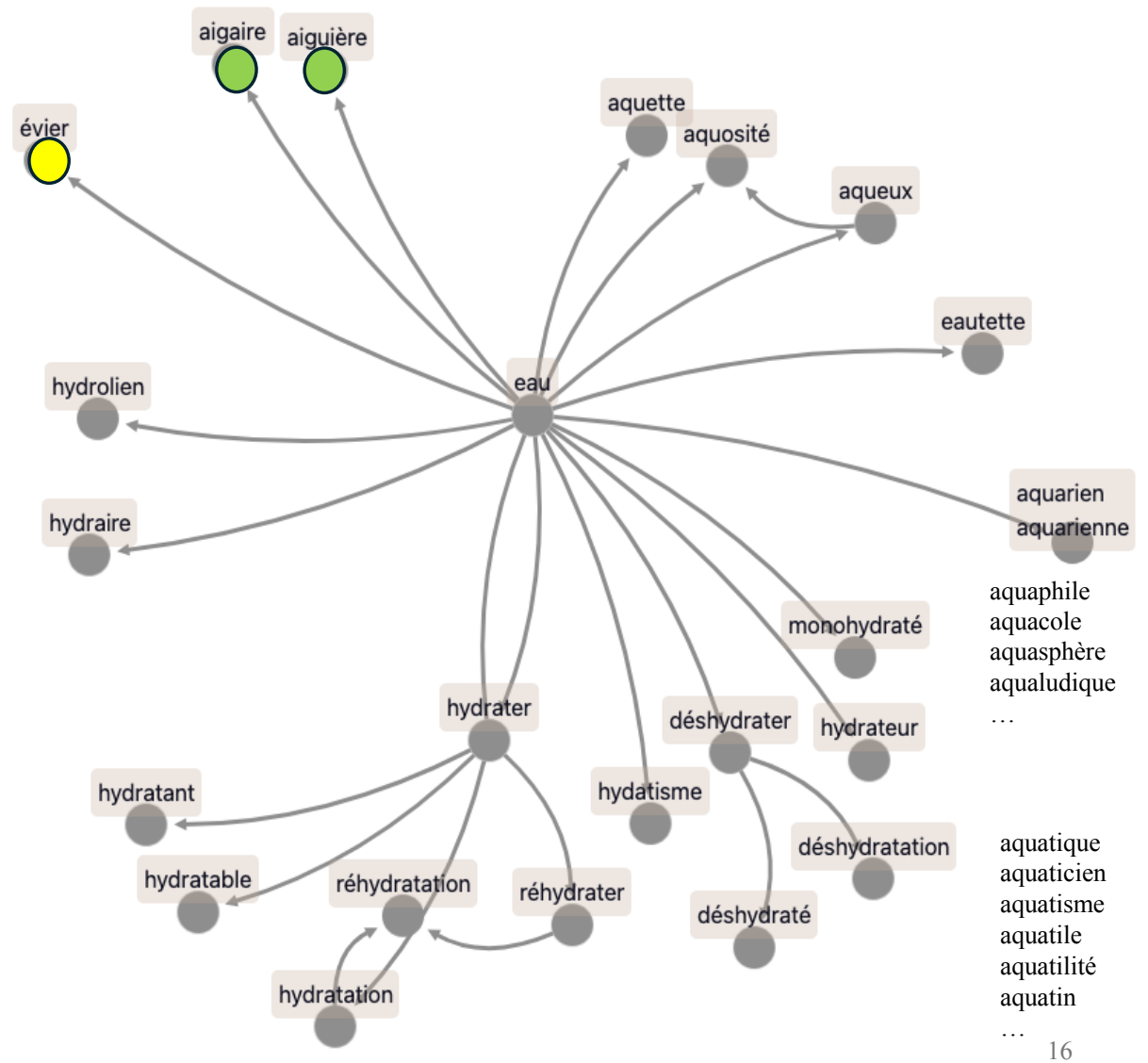
# Maximal distance : stem suppletion

- Stem integrity (C2) and pattern frequency (C3) do not apply
- Family distribution?
- Example with *eau* 'water'



# Maximal distance : criteria 1-3 fail

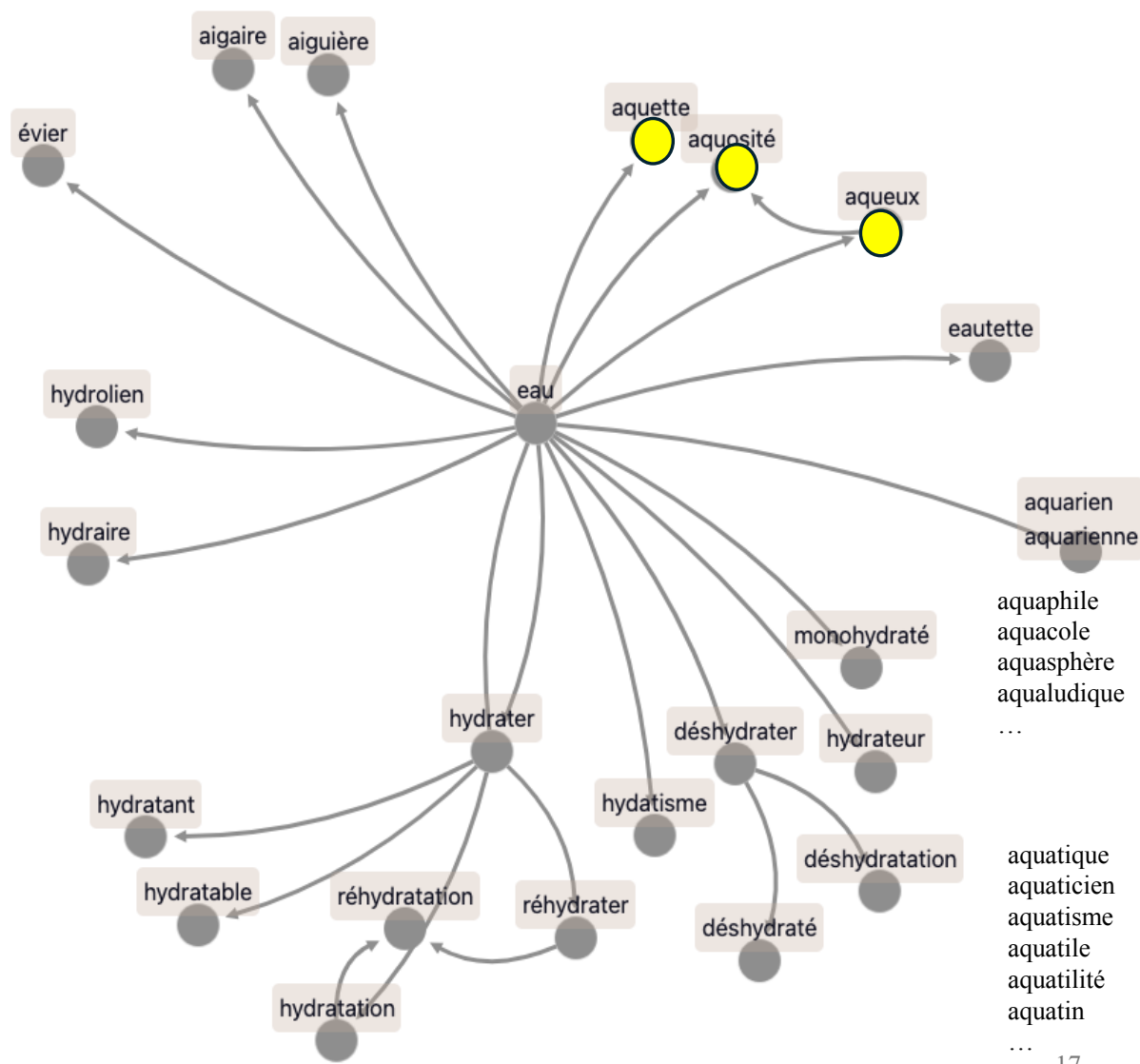
- Family distribution:
- [ev]
- [eg]





# Maximal distance : criteria 1-3 fail

- Family distribution:
- [ev]
- [eg]
- [ak]



# Maximal distance : criteria 1-3 fail

- Family distribution:

- [ev]

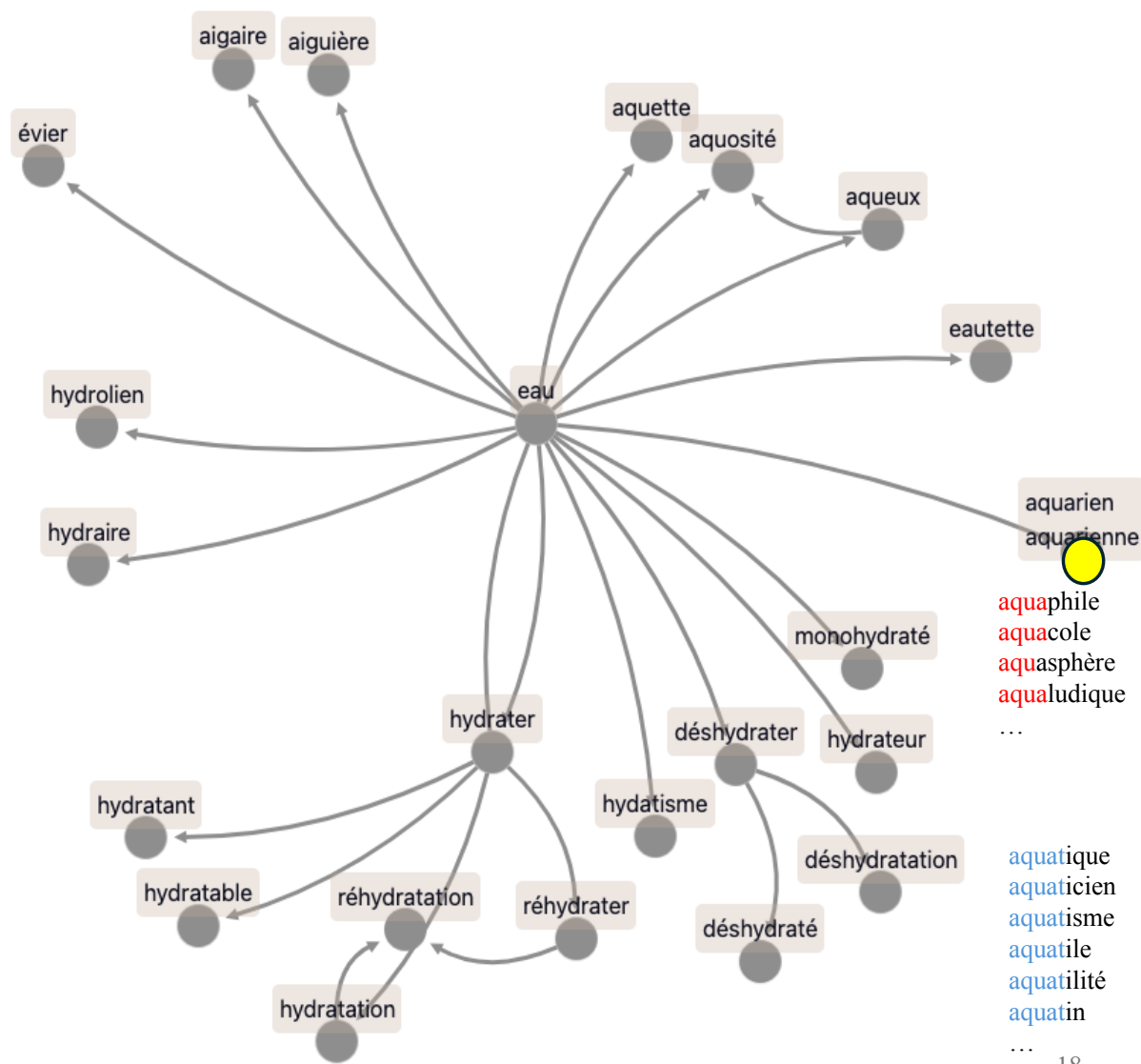
- [eg]

- [ak]

- [akwar]

- [akwa]

- [akwat]



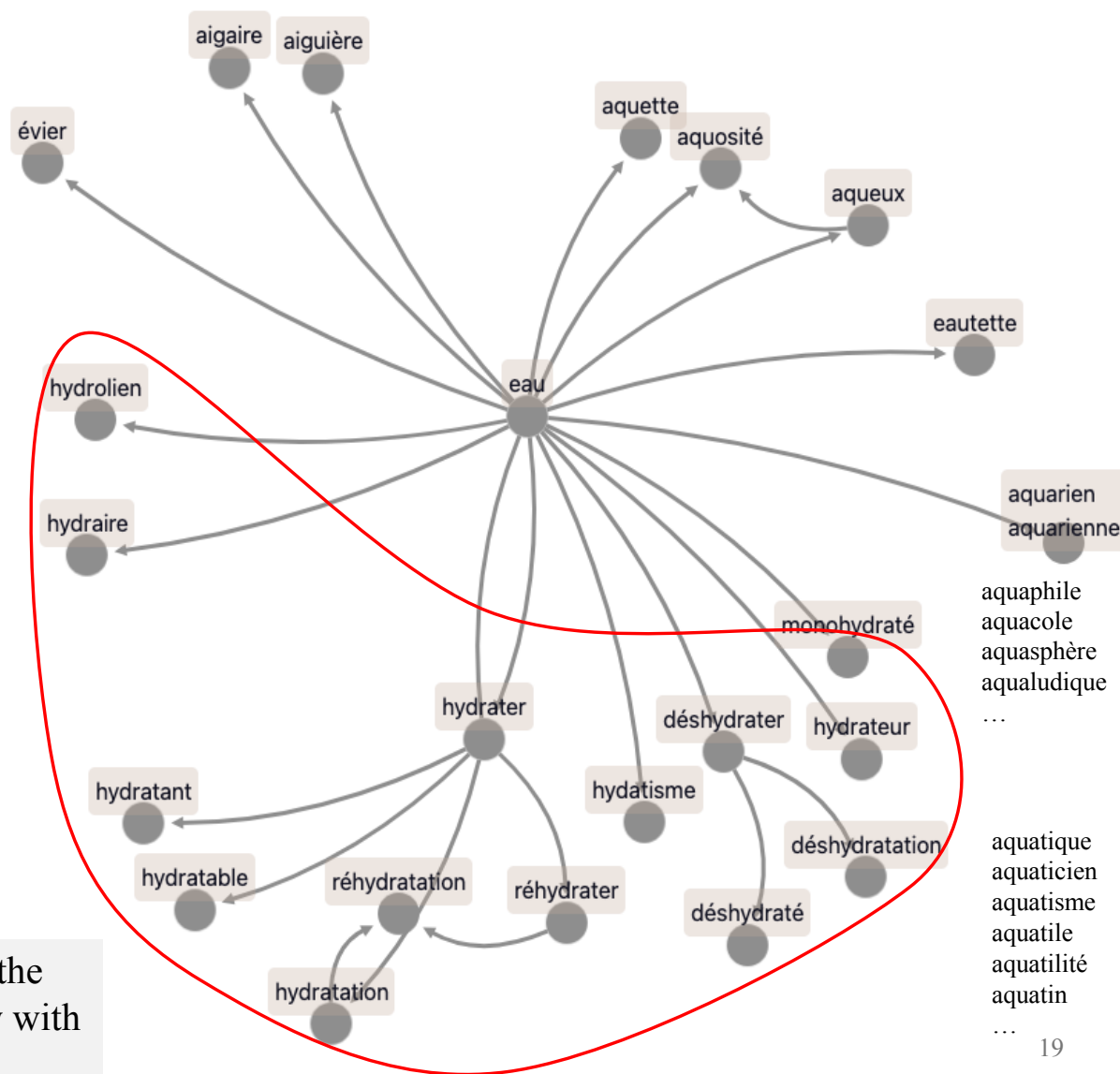
# Maximal distance : criteria 1-3 fail

• Family distribution:

- [ev]
- [eg]
- [ak]
- [akwar]
- [akwa]
- [akwat]
- [idr]



stem distribution in the family is the reflect of the speakers' familiarity with each stem.



# "Intermediary" formal distance =4

<i>grain</i> [grɛ̃] / <i>granul-é</i> [granyl] 'grain' / 'granule'	C4
C2=yes	<i>granulé</i>
C3= no	<i>granuler</i> <i>granulaire</i> <i>granularité</i> ...
<i>proche</i> [prɔʃ] / <i>proxim-ité</i> [prɔksim] 'near' / 'proximity'	<i>proximal</i>
C2=yes	<i>approximer</i>
C3= no	<i>approximation</i> <i>approximatoire</i> ...
	C4

# Minimal formal distance (C1=1)

- *abricot/abricotier* 'apricot' / 'apricot tree',
  - *champ /champêtre* 'field' / 'rural',
  - *clou/clouter* 'nail' / 'nail<sub>v</sub>',
  - *nuit/nuisette* 'night' / 'babydoll',
- 
- Often, epenthesis analysis draws on etymological, phonological motivations, or on the grapheme-phoneme correspondence.

Pagliano, 2004  
Plénat & Roché, 2012

- Corpus-based complementary explanation, using C3 and C4?

	<b>C3</b>	<b>C4</b>
<i>abricot/abricot-ier</i> [abriko] / [abrikot]	54% of consonant epentheses are [t]	(maximal distribution of alternant stem)
<i>clou/clout-er</i> [klu/klut]		<i>cloutier, cloutage, cloutière, clouté</i>
<i>champ/ champ-être</i> [ã] / [ãp]	29% of consonant epentheses realize the latent consonant and are ≠ from [t]. Here: <i>drap/drapier, camp/camper</i>	(maximal distribution)
<i>nuit/nuis-ette</i> [nɥi]/[nɥiz]	Other patterns (17%). Here, only: <i>rire/risette</i>	no

# Conclusion

- Our assumption: variation is (also) about paradigms
- Derivational resources structured into families and series: Phononette.
- In addition to criteria based on historical factors / focusing on strict string comparison:
  - The impact of morphophonological series / classes: frequency of each type variation among derivational relations
  - The role of word families: propagation of each derivational stem among family words
- Promising results, showing that the question of formal variation should not be considered (only) at phonemic level, but that it also benefits from the global level of derivation (provided a paradigm-based approach is used as a model for the database).

# Conclusion

- We haven't look at all the phenomena (apophony, assibilation, haplology...)
- The Phononette portion we worked on is a sandbox
  - Quantitative results are at best trends
  - The presented criteria must be applied to the complete Phononette coverage
- Quality of the morphophonological coding has to be fully checked
  - stem/affix boundary
- String distance algorithms more elaborate than Levenshtein must be tested
- Size of families must be included in C4, as token frequency (in addition to type frequency) for C3 and C4
- Weight criteria? Psycholinguistic experiments?



# Bibliography

- Bonami, O & Boyé, G. (2003). "Supplétion et classes flexionnelles dans la conjugaison du français". *Langages* **152**:102–126.
- Bonami, O & Boyé, G. (2005). « Construire le paradigme d'un adjectif ». *Recherches linguistiques de Vincennes* **34**:77–98.
- Boyé, G. (2006). "Suppletion". In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, tm. 12, p. 297-299. Oxford : Elsevier.
- Carstairs, A. (1988) "Some implications of phonologically conditioned suppletion" in G. Booij, J. van Marle (eds), *Yearbook of Morphology*. Dordrecht: Kluwer, 67-94.
- Corbin, D. (1985), "Les bases non autonomes en français ou comment intégrer l'exception dans un modèle lexical." *Langue française* **66**, 54–76.
- Corbin, D. (1987), *Morphologie dérivationnelle et structuration du lexique*. 2 volumes. Tübingen : Niemeyer.
- Dressler, W.U. (1985) "Suppletion in word formation" in J. Fisiak (ed.), *Historical Semantics. Historical Word-Formation*. Berlin – New York: Mouton, 97-112.
- Dressler, Wolfgang U.. (2015) "27. Allomorphy". *Volume 1 Word-Formation: An International Handbook of the Languages of Europe*, Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen and Franz Rainer (eds.), Berlin, München, Boston: De Gruyter Mouton. 500-516.
- Frei, H. (1929). *La grammaire des fautes : introduction à la linguistique fonctionnelle*. Paris : Geuthner – Genève : Kundig – Leipzig: Harrassowitz
- Hathout, N., Sajous F. & Calderone B. (2014). "GLÀFF, a Large Versatile French Lexicon". *Proceedings of LREC'14*, Reykjavik, Iceland. 1007-1012
- Kiparsky, P. (1996) "Allomorphy or morphophonology?" in R. Singh (ed.), Trubetzkoy's Orphan. *Proceedings of the Montréal Roundtable « Morphology: Contemporary Responses »*. Amsterdam: John Benjamins, 13-31.
- Mel'čuk, I.A. (2006) *Aspects of the Theory of Morphology*. Berlin – New York: Mouton de Gruyter.
- Namer, F., et al. (2023). "Démonette-2, a derivational database for French with broad lexical coverage and fine-grained morphological descriptions." *Lexique* **33**: 6-40.
- Pagliano, C. (2004) "Elaboration d'un corpus morphophonologique : l'épenthèse consonantique à la frontière suffixale en français", *Corpus* **3**
- Plénat, M. (2008) "Le thème L de l'adjectif et du nom", *CMLF*. Paris: ILF, 1613-1626
- Plénat, M & Roché M (2012) "Elements de morphologie quotidienne. Contribution à l'étude des conditionnements de l'épenthèse présuffixale". *CMLF* Paris: ILF, 1369-1386
- Tranel, B. (1981) *Concreteness in Generative Phonology. Evidence from French*. Berkeley – Los Angeles: University of California Press.