

Affixal rivalry in demonym formation

Juliette Thuilier*, Delphine Tribout•, Marine Wauquier°

*U. Toulouse Jean Jaurès & CLLE, •U. Lille & STL, °U. Paris & LLF

Workshop on affix rivalry, March 19, 2021

Introduction

- Rival processes are often defined as processes that operate on the same type of bases and produce the same type of lexemes, but differ on the phonological level (Plag 1999).
- We chose to study rivalry in French demonym formation because:
 - ▶ there are many processes forming demonyms in French
 - ▶ the domain is easily delimited
 - ▶ the semantic part of all processes can easily be compared
 - 🗨️ we define the semantics of all demonyms as 'inhabitant of'
- Goals:
 - ▶ to study the affixal formation of demonyms in French
 - ▶ to test some hypotheses of (Roché & Plénat 2016)
 - ▶ to propose a statistical modelling of rival processes forming demonyms
- Question:
 - ▶ are there properties that allow to predict the choice of the suffix?

Outline

- 1 Data
- 2 Local observations
- 3 Global observations
- 4 Conclusion

Data

- We used the **Prolex Database**, a multilingual dictionary of proper names, which also contains information about relational names and adjectives associated to proper nouns
- We extracted a list of French toponyms with their demonyms (adjectives or nouns)
 - 👉 10,213 paires
- We got phonetic transcription (with syllabation) of some toponyms and demonyms from the French Wiktionary
 - ▶ we only kept pairs for which we had at least the toponym's transcription
 - 👉 2,702 pairs

Data

Annotation

- All pairs were automatically or manually annotated according to different criteria:
- **Morphological** criteria:
 - ▶ Is the toponym a compound?
 - ▶ How is the demonym formed (conversion or suffixation)?
 - ▶ Is the demonym derived from an allomorphic base of the toponym?
- **Phonological** criteria
 - ▶ Number of segments in the toponym
 - ▶ Number of syllables in the toponym
 - ▶ Last segment of the toponym
 - ▶ Last consonnant of the toponym
 - ▶ Last vowel of the toponym
 - ▶ Ratio of nasal vowels (number of nasal V / total number of V) in the toponym
 - ▶ Backness score of the vowels in the toponym (front V coded 1 and back vowel coded 3) (following Lohmann 2017)

Data

- We ended up with a lot of suffixes having very few occurrences
- We decided to limit the scope of the study to the 4 more productive suffixes: *-ais*, *-éen*, *-ien*, *-ois*
 - 📖 2,218 paires
- Suffix distribution

suffix	#	%
ais	555	25.02
éen	165	7.44
ien	644	29.04
ois	854	38.50
total	2218	100.00

NB: *-en* was considered as an allomorph

- ▶ of *-éen* when the toponym ends in [e] or [ɛ] (Vendée→vendéen)
- ▶ of *-ien* when the toponym ends in [i] (Algérie→algérien)

Data

Compounds

- 611 toponyms are compounds (27,5% of the data).
- Different cases in the formation of demonyms out of compounds toponyms:
 - ▶ 271 derive from the 1st element of the toponym (**Dives**-sur-Mer→**divais**)
 - ▶ 114 derive from the whole toponym (**Lot-et-Garonne**→**lot-et-garonnais**)
 - ▶ 96 derive from the last element of toponym (Saint-**Tropez**→**tropézien**)
 - ▶ 130 other complexe cases (**Cinq-Mars-La-Pile**→**cinq-marsien**, Saint-Omer→audomarois)
- The issue of what forms the radical of demonyms deriving from compounds has not been addressed yet.
- 👉 For the phonological properties of compound toponyms, we only took into account what is used as radical in the demonym

Data

Examples

- ais* Antilles→antillais, Bayonne→bayonnais, Népal→népalais
Bagneux→bagnolais, New York→new-yorkais
- éen* Guadeloupe→Guadeloupéen, Guinée→guinéen, Nancy→nancéen
Foix→fuxéen, Noisy-le-Sec→noiséen
- ien* Nanterre→nanterrien, Algérie→algérien, Athènes→athénien
Sochaux→sochalien, Saint-Maurice→saint-mauricien
- ois* Belleville→bellevillois, Strasbourg→strasbourgeois, Turin→turinois
Meaux→meldois, Le Blanc-Mesnil→blanc-mesnilois

Outline

- 1 Data
- 2 Local observations**
- 3 Global observations
- 4 Conclusion

Number of syllables

Syll.	-ais		-éen		-ien		-ois		Total
	#	%	#	%	#	%	#	%	
1	63	16.36	14	3.64	71	18.44	237	61.56	385
2	312	26.80	104	8.94	349	29.98	399	34.28	1164
3	148	26.86	37	6.71	179	32.49	187	33.94	551
4+	32	27.12	10	8.47	45	38.14	31	26.27	118

- No clear effect, but 2 tendencies

- ▶ Monosyllabic bases favor **-ois**

ex. Lille → lill**ois** Cannes → canno**is**
[lil] [kan]

- ▶ Long bases (≥ 4 syll) favor **-ien**

ex. Mésopotamie → mésopotam**ien** Saint-Léonard → saint-leonard**ien**
[me.zo.po.ta.mi] [sɛ̃.le.o.naʁ]

- Multinomial regression (with 10-fold cross-validation) shows poor results

- ▶ Average accuracy: 0.3877

Final consonant vs. final vowel

	-ais		-éen		-ien		-ois	
	#	%	#	%	#	%	#	%
C	259	23.52	14	1.27	247	22.43	581	52.77
V	296	26.50	151	13.52	397	35.54	273	24.44

- Final consonants favor **-ois**

ex. Lille → lillo**ois** Cannes → canno**is** Aurillac → aurillaco**is**
 [lil] [kan] [oʁijak]

	-ais		-éen		-ien		-ois	
	#	%	#	%	#	%	#	%
Front V	69	8.97	136	17.69	358	46.55	206	26.79
Other	486	33.54	29	2.00	286	19.74	648	44.72

- Front vowels favor **-ien**

▶ Both **-ien** and **-éen** have a preference for bases ending with a front vowel

ex. Ivry → ivri**en** Vertus → vertus**ien** Vendée → vend**éen** Sanxay → sanx**éen**
 [ivʁi] [vʁɛty] [vãde] [sãse]

Backness as a predictor

- From the backness of the last segment (front vowel) to the overall backness (backness score)
- To some extent a relevant feature to discriminate the suffixes
 - ▶ Significant feature but low accuracy (0.3891) for all 4 suffixes
 - ▶ Moderate accuracy (0.6352574) for the prediction of *-éen* and *-ien* vs. *-ais* and *-ois*
 - ▶ Moderate accuracy (0.6146379) for *-ais* and *-ois*

Properties of the last segment

	-ais		-éen		-ien		-ois	
	#	%	#	%	#	%	#	%
plosive C	56	32.00	7	4.00	29	16.57	83	47.43
approx. C	75	27.47	3	1.1	29	10.62	166	60.81
alv. fric.	16	12.03	1	0.75	66	49.63	50	37.59
other fric.	39	11.82	3	0.91	98	29.7	190	57.58
other seg.	369	28.23	151	11.55	422	32.29	365	27.93

- Plosive and approximant consonants favor *-ois*

ex. Dunkerque → dunkerqu**ois** Étampes → étamp**ois** Sète → séto**is** Lille → lillo**is**
 [dœkɛʁk] [etɑ̃p] [set] [li]

- Fricatives favor *-ois* and *-ien*

- ▶ Alveolar fricatives (z, s) ⇨ *-ien*

ex. Alsace → alsac**ien** Mulhouse → mulhous**ien**
 [alzas] [myluz]

- ▶ Other fricatives (v, f, ʒ, ʃ, ʁ) ⇨ *-ois*

ex. Loches → loch**ois** Orange → orang**ois** Quimper → quimpéro**is**
 [lɔʃ] [orɑ̃ʒ] [kɛ̃pɛʁ]

Dissimilative constraint

- A base with a final nasal segment (V ou C) is unlikely to combine with a suffix containing a nasal vowel (-*ien* and -*éen*)

	-ais		-éen		-ien		-ois	
	#	%	#	%	#	%	#	%
nasal	292	53.68	3	0.55	40	7.35	209	38.42
vow	219	61.86	3	0.85	15	4.24	117	33.05
cons	73	38.42	0	0	25	13.16	92	48.42
other	264	15.77	162	9.68	603	36.02	645	38.53

- Nasal vowels favor largely **-ais**, especially back nasal vowels (206/219)

ex. Avignon → avignon**ais** Châlons-en-Champagne → chalon**ais**
 [aviɲɔ̃] [ʃalɔ̃]

ex. Orléans → orléan**ais** Villers-Allerand → villers-allerand**ais**
 [ɔʁleɑ̃] [vilɛʁalɛʁɑ̃]

Discrimination of *-ais* and *-ois*

- Probability to have *-ais* instead of *-ois* based on the presence of a final consonant, the backness score and the ratio of nasal vowels of the base
 - ▶ Generalized Linear Model with 10-fold cross-validation
 - ▶ Accuracy of 0.651

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.2180743	0.17449425	-6.980599	2.939246e-12
Final consonant	-0.7832557	0.12428535	-6.302076	2.936852e-10
Backness score	0.7097344	0.08602953	8.249893	c1.585358e-16
Nasal vowel ratio	-0.1636630	0.21845175	-0.749195	4.537397e-01

- Redundancy between backness score and nasality
 - ▶ Correlation of 0.425
- Similar accuracy when alternating the predictors

Outline

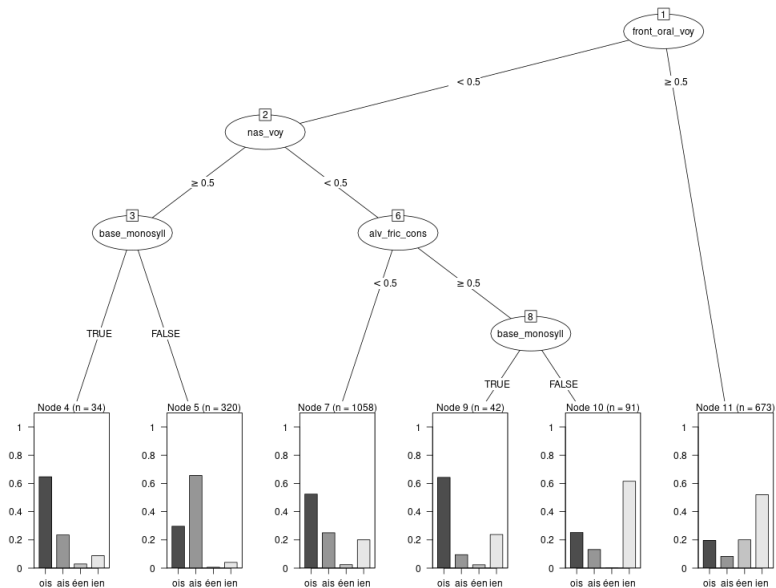
- 1 Data
- 2 Local observations
- 3 Global observations**
- 4 Conclusion

Multinomial regression

- Probability to have *-ais*, *-éen* or *-ien* instead of *-ois*
 - ▶ Accuracy: 0.558

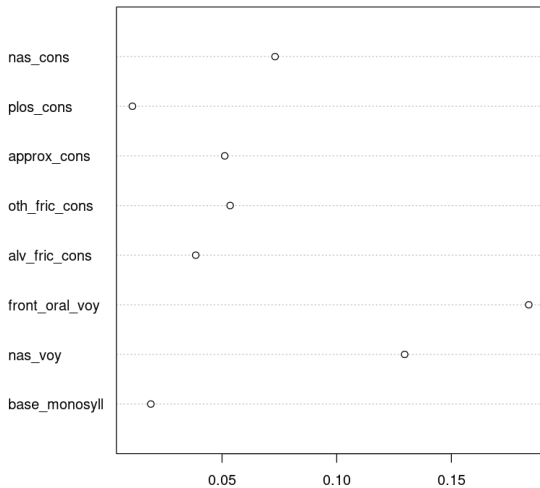
	<i>-ais</i>	<i>-éen</i>	<i>-ien</i>
MONOSYLLABIC	-0.981*** (0.163)	-0.530 (0.321)	-0.865*** (0.162)
ALV FRIC	-0.934 (0.416)	-3.228* (1.068)	0.129 (0.335)
OTHER FRIC	-1.433*** (0.347)	-3.496*** (0.677)	-0.860** (0.302)
APPROX C	-0.757* (0.330)	-3.430*** (0.678)	-2.046*** (0.341)
NASAL C	-0.046 (0.339)	-10.058 (22.299)	-1.469*** (0.357)
PLOS C	-0.234 (0.347)	-1.816*** (0.524)	-11.242*** (0.351)
FRONT V	-0.910** (0.339)	0.548 (0.369)	0.591* (0.294)
NASAL V	0.612 (0.320)	-3.111*** (0.680)	-2.334*** (0.384)
INTERCEPT	0.113 (0.300)	-0.486 (0.352)	0.440 (0.277)

Conditional Inference Tree



Random Forest

Variable importance



Outline

- 1 Data
- 2 Local observations
- 3 Global observations
- 4 Conclusion**

Conclusion

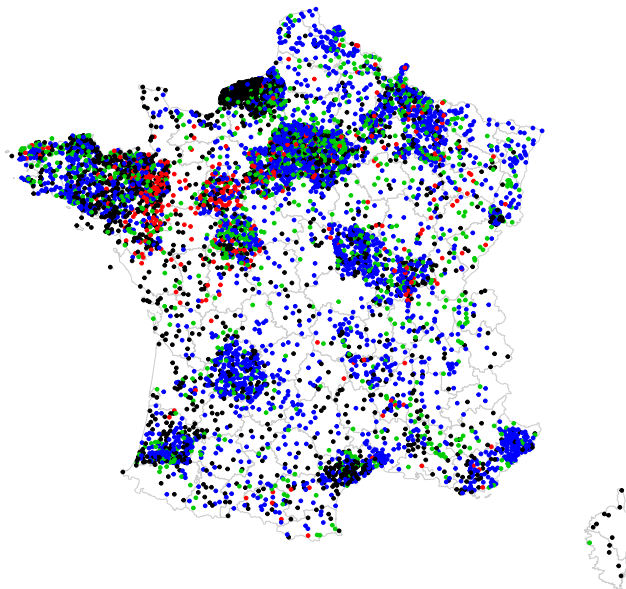
- No clear discriminative criteria, only tendencies.
- **Monosyllabic bases** favour *-ois*, but their number is small (17%)
- Final **front vowels** favor *-éen* and *-ien* as opposed to *-ais* and *-ois*. Then they favor *-ois* over *-ais*
 - ☞ it confirms Roché & Plénat's observation that front vowels are under-represented before *-ais* and over-represented before *-ois* as opposed to other suffixes (they did not study *-ien* and *-éen*)
- Final **nasal segments** (consonant or vowel) favor *-ais*. They also clearly disadvantage *-ien* and *-éen*.
 - ☞ it is a dissimilative constraint, like those found by Roché & Plénat
- Final **alveolar fricatives** favor *-éen* and *-ien*
 - ☞ it confirms Roché & Plénat's observation that [s] and [z] are uncommon before *-ais* and *-ois* because of a dissimilative constraint with the feminine form of both suffixes (*-aise* [ɛz] and *-oise* [waz])

Future Work

- We will explore other criteria
- Other **phonological properties**
 - ▶ Dissimilative constraint between the suffix and the last vowel of the toponym
- **Morphological analysis** of toponyms
 - ▶ Correlation between the suffix and the elements used as radical for the formation of the demonym
 - ▶ Correlation between the suffix and some endings, such as *-bourg* (Stras**bourg**), *-ville* (e.g. Deau**ville**), *-mont* (e.g. Cler**mont**)
- **Geographical distribution** of the suffix
 - ▶ Correlation between the suffix and the geographical location (in France)
 - ▶ Impact of geographical neighbours on the choice of the suffix (in France)

Geographical distribution of the suffixes

- 6517 toponyms



Thank you for your attention!

This study was supported by the *Demonext* research project, which is funded by the French National Research Agency (ANR) and identified as ANR-17-CE23-0005



References

- Lohmann, A. 2017. Phonological properties of word classes and directionality in conversion. *Word Structure* 10(2), 204-234
- Plag, I. 1999. *Morphological productivity*. Berlin: Mouton de Gruyter.
- Roché, M. & M. Plénat. 2016. De l'harmonie dans la construction des mots français. *Actes du CMLF 2016*.