

French ethnonyms, toponyms, demonyms and their paradigmatic organization

Nabil Hathout^a, Fiammetta Namer^b, Michel Roché^c

^aCLLE, CNRS & Université de Toulouse Jean Jaurès

^bATILF, Université de Lorraine & CNRS

^cUniversité de Toulouse Jean Jaurès

ParadigMo 2

3 June 2021



Outline

- 1 Introduction
- 2 Semantic content
- 3 Formal content
- 4 ParaDis
- 5 Conclusion

Roché (2008, 2017); Strnadová (2014); Schalchli & Boyé (2018) argue that the morphological families of French names of inhabitants are organized paradigmatically.

- 1 What are the paradigms defined by these families?
- 2 What generalizations emerge from their paradigmatic description?
- 3 How could we describe them in ParaDis?
(Hathout & Namer, 2018; Namer & Hathout, 2020)

Contributions

- The names of inhabitants do not belong to a single homogeneous paradigm.
- A confirmation that the semantic and formal properties of families of the names of inhabitants are largely disconnected.
- With some exceptions, the families of the names of inhabitants do not contain at the same times a name of an ethnic group and a name of a town.
- The introduction of morphological constraints in ParaDis.

Terminology

demonym = name of people from a particular geographical entity, town (td) (*parisien*_{td} 'Parisian') or country (ct) (*français*_{cd} 'French person')

- il a discuté avec un parisien 'he discussed with a Parisian'
- il a discuté avec un français 'he discussed with a Frenchman'

ethnonym = name of humans who belong to a people or an ethnic group (ed), regardless of any territorial anchorage (*touareg*_{ed})

- il a discuté avec un touareg 'he discussed with a Tuareg'

toponym = name of towns (tt) and names of countries (ct) (*Paris*_{tt}; *France*_{ct})

- il travaille à Paris 'he works in Paris'
- il travaille en France 'he works in France'

Terminology

relational adjectives (ra) = adjectives that correspond to all these nouns in French (*parisien*_{ra-td,ra-tt}; *français*_{ra-cd,ra-ct}; *touareg*_{ra-ed})

- l'élégance parisienne 'Parisian elegance'
- les arrondissements parisiens 'the Parisian districts'
- l'élégance française 'French elegance'
- les régions françaises 'the French regions'
- l'élégance touareg 'Tuareg elegance'

The families may also contain names of **languages** and their relational adjectives but we will not consider them in this study.

- il parle français 'he speaks French'
- la grammaire française 'French grammar'

The cells of the paradigms

The paradigm of a demonym may have up to 10 cells.

	ethnic group	town	country
toponym		tt : name of town	ct : name of country
toponym RA		ra-tt : RA of town toponym	ra-ct : RA of country toponym
demonym	ed : ethnic group name	td : name of town inhabitant	cd : name of country inhabitant
demonym RA	ra-ed : RA of ethnonym	ra-td : RA of town demonym	ra-cd : RA of country demonym

Systematic syncretism

The members of the families of the demonyms cannot be described by means of WFR because of the many syncretisms they display.

ed/ra-ed Ethnonyms and their relational adjectives have the same form.

td/ra-td/ra-tt Town demonyms, their relational adjectives and the relational adjectives of the corresponding town toponyms have the same form.

cd/ra-cd/ra-ct Country demonyms, their relational adjectives and the relational adjectives of the corresponding country toponyms have the same form.

	ethnic	town	country
toponym		(tt) Bordeaux	(ct) France
toponym RA		(ra-tt) bordelais	(ra-ct) français
demonym	(ed) ouzbek	(td) bordelais	(cd) français
demonym RA	(ra-ed) ouzbek	(ra-td) bordelais	(ra-cd) français

Systematic syncretism

- Because **ra-tt** is derived from **tt**, WFR cannot account for its syncretism with **ra-td**. The same goes for **ra-ct**.
- Paradigms can account for these syncretisms (Bochner, 1993).

Bordeaux_{tt}, bordelais_{ra-tt}, bordelais_{td}, bordelais_{ra-td}

$$\left\{ \left[\begin{array}{c} /X/ \\ \text{Np} \\ \text{'Z'} \end{array} \right], \left[\begin{array}{c} /X-ais/ (= /Y/) \\ \text{A} \\ \text{'of Z'} \end{array} \right], \left[\begin{array}{c} /Y/ \\ \text{N} \\ \text{'inhabitant of Z' (= 'T')} \end{array} \right], \left[\begin{array}{c} /Y/ \\ \text{A} \\ \text{'of T'} \end{array} \right] \right\}$$

Simplification

The relational adjectives can be fully predicted semantically and formally from the meaning and form of the nouns included in the paradigm:

- A relational adjective of a noun has the same meaning as the noun
- The relational adjective of an ethnonym has the same form as the ethnonym.
- The relational adjective of an demonym has the same form as the demonym.
- The relational adjective of a toponym has the same form as its demonym.

The paradigm is fully determined by the morphological relations that exist between the nouns

ed	tt	td	ct	cd

More syncretism

ouzbek

- $ouzbek_{cd}$ is the demonym of $Ouzbekistan_{ct}$.
- Formally, $Ouzbekistan_{ct} \not\rightarrow ouzbek_{cd}$ ($France_{ct} \rightarrow franais_{cd}$).
- $ouzbek_{cd}$ is a **copy** of $ouzbek_{ed}$

The same goes for *turc*, *tchétchène*, *russe*, *afghan*, *birman*, *papou*, etc.

- $ouzbek_{cd}$ is derivative of $Ouzbekistan_{ct}$ but WFR cannot account for the construction of its form.
 - WFR cannot account for the syncretism between $ouzbek_{cd}$ and $ouzbek_{ed}$.
- **Derivational paradigms** can account for both.

Outline

- 1 Introduction
- 2 Semantic content**
- 3 Formal content
- 4 ParaDis
- 5 Conclusion

2-cells semantic paradigms

Concepts

E ethnic group;

T town; **TI** town inhabitant;

C country; **CI** country inhabitant

	E	T	TI	C	CI
S1		Paris Moscou	parisien moscovite		
S2				Bolivie Moldavie France	bolivien moldave français

S1: T
↓
TI

S2: C
↓
CI

2-cells semantic paradigms

Semantic constraints

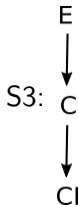
- A town has inhabitants: $\forall \mathcal{S}, \exists X_T \in \mathcal{S} \Rightarrow \exists Y_{TI} \in \mathcal{S}$
- A country has inhabitants: $\forall \mathcal{S}, \exists X_C \in \mathcal{S} \Rightarrow \exists Y_{CI} \in \mathcal{S}$

Semantic relations

- **TI** are the inhabitants of **T**
- **CI** are the inhabitants of **C**

3-cells semantic paradigm

	E	T	TI	C	CI
S3	kurde russe corse			Kurdistan Russie Corse	kurde russe corse

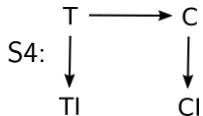


Semantic relations

- **C** is the country (originally) populated by **E**
- **CI** are the inhabitants of **C**

4-cells semantic paradigm

	E	T	TI	C	CI
S4		Alger Luxembourg Toul Bordeaux	algérois luxembourgeois toulois bordelais	Algérie Luxembourg (le) Toulois (le) Bordelais	algérien luxembourgeois toulois bordelais



Semantic relations

- **T** is the capital or the metropolis of **C**
- **TI** are the inhabitants of **T**
- **CI** are the inhabitants of **C**

5-cells semantic paradigm are exceptions

	E	T	TI	C	CI
	vénète	Venise	vénitien	Vénétie	vénète
	outaouais	Ottawa	ottavien	Outaouais	outaouais

- Complete families are exceptions.
- Their paradigms cannot be complemented by new families.
- The semantic relation between the ethnic group and the country / the city is only relevant from a diachronic point of view.
- The orientation of the semantic relation between the country and the city is unclear.

Semantic organization: summary

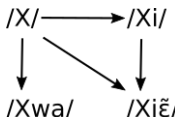
- With few exceptions, semantic paradigms contain either an ethnic group or a city.
- Systematic prediction:
 - a town has inhabitants
 - a country has inhabitants.
- The cells of all the semantic paradigms are indexed by the same categories.

Outline

- 1 Introduction
- 2 Semantic content
- 3 Formal content**
- 4 ParaDis
- 5 Conclusion

Size of the formal paradigms

- The formal and the semantic contents of the paradigms of the demonyms are largely disconnected.
- The number of different forms included in a formal paradigm does not predict the size of the semantic paradigm

1	corse	/X/
2	ouzbek, Ouzbékistan Bordeaux, bordelais France, français	/X/ → /Xistã/ /X/ → /Xε/
3	thai, Thaïlande, thaïlandais outaouais, Ottawa, ottavien	/X/ → /Xlãd/ → /Xlãde/
4	Alger, algérois, Algérie, algérien Venise, vénitien, Vénétie, vénète	

Exponents

The paradigms of the demonyms are formally very diverse.

	X	Xien	Xain	Xais	Xois	Xie	Xistan
de	ouzbek kurde slovaque	phénicien					
tt	Bordeaux Paris Toul						
td		parisien	toulousain	agenais	algérois toulois		
ct	Prusse			Lyonnais Bordelais	Toulois	Algérie Turquie	Kurdistan Ouzbékistan Afghanistan
cd		tunisien prussien	cubain marocain	ecossais irlandais	bénois bavarois toulois		

Exponents

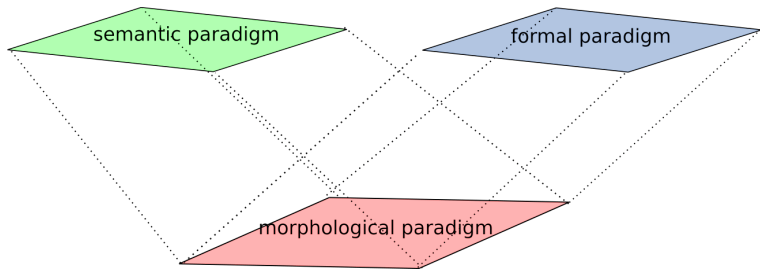
- $X \rightarrow Xie$ and $X \rightarrow Xistan$ are only used to derive country toponyms.
 - $X/Xie \rightarrow Xien$ is only used to derive demonyms (country or city)
 - $X \rightarrow Xais$, $X \rightarrow Xoïs$ and $X \rightarrow Xain$ are used to coin demonyms (country or city) and country toponyms.
 - Simplex nouns are either ethnonyms or town toponyms.
 - *Prusse* and *phénicien* are exceptions.
- The formal and semantic paradigms of a morphological family are **not predictable** from one another

Outline

- 1 Introduction
- 2 Semantic content
- 3 Formal content
- 4 ParaDis**
- 5 Conclusion

General Principles

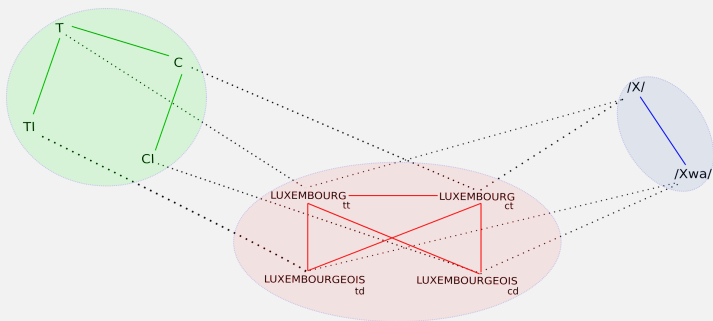
- ParaDis: a paradigm-based model
- The notion of paradigm is generalized (meaning, form, category)
- Formal and semantic paradigms are independent and unconnected
- The formal and semantic paradigms combine into a morphological paradigm



Correspondences with the morphological level

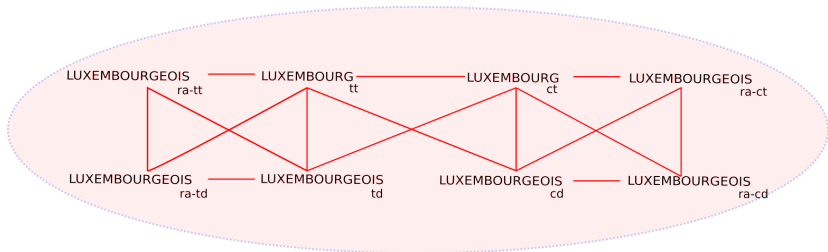
- Different graphs at semantic and formal levels
- The morphological paradigm is “shaped” by the correspondences
- tt/ct and td/cd syncretism is described by the correspondences between the formal and the morphological paradigms.

Luxembourg_{tt}



Relational adjectives

- The “core” network is complemented by the RAs of the noun in the morphological paradigm
- RAs are involved in systematic syncretic relations
 - ed = ra-ed
 - td = ra-td = ra-tt
 - cd = ra-cd = ra-ct



Relational adjectives

The syncretic relations hold in **any morphological paradigm** that contains an **ed**, **tt**, **td**, **ct** or **cd** lexeme. They are bound to the lexical nature of ethnonyms, toponyms and demonyms.

- In existing paradigms, they is enforced by **the correspondences**.
- A set of **morphological constraints** that ensure that the syncretic relations also hold in any new paradigm.

td = ra-td

For any morphological paradigm \mathcal{M} ,

if \mathcal{M} contains a noun L in correspondence with a TI H (at the semantic level) and with a form D (at the formal level),

then \mathcal{M} also contain an adjective L' in correspondence with H (at the semantic level) and with D (at the formal level).

Superposition of morphological paradigms

Morphological paradigms can be superposed in order to account for the intuition that they are defined on the same set of semantic relations and that they present the same syncretic regularities.

The superposition is a single **derivational paradigm** made up of several different morphological paradigms.

ed	tt	td	ct	cd
	Paris	parisien	France	français
russe			Russie	russe
	Alger	algérois	Algérie	algérien
vénète	Venise	vénitien	Vénétie	vénète

- The morphological families of all the morphological paradigms fit in the same derivational paradigm.

Outline

- 1 Introduction
- 2 Semantic content
- 3 Formal content
- 4 ParaDis
- 5 Conclusion**

Conclusion

- Ethnonyms, demonyms and toponyms are a textbook case for paradigmatic morphology
- Semantic and formal regularities form **distinct paradigms**
- The apparent referential unity of these data is misleading with **several semantic abstract paradigms** “rooted” in an **E**, or a **T**, or a **C**

- Formal constraints apply to the formal paradigms
(eg. hiatus avoidance: *bordelais* ← *Bordeaux*)
- Semantic constraints apply to the semantic paradigms
(eg. a town has inhabitants)

Conclusion

- Syncretic relations are enforced by the correspondences in the existing paradigms and by morphological constraints.
- Morphological constraints apply at the morphological level: they cannot be reduced to purely formal constraints nor purely semantic constraints
- Some toponyms and demonyms have several RAs:
kurde / kurdistanais, croate / croatien, basque / basquais.
- In ParaDis, n -uplets are accounted for as a superposition of different morphological families that are in correspondence with the same semantic family.

Références

- Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin & New-York: Mouton de Gruyter.
- Hathout, Nabil & Fiammetta Namer. 2018. La parasyntèse à travers les modèles : des RCL au ParaDis. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 365–399. Langage Sciences Press.
- Namer, Fiammetta & Nabil Hathout. 2020. ParaDis and Démonette – from theory to resources for derivational paradigms. *The Prague Bulletin of Mathematical Linguistics* 114. 5–33.
- Roché, Michel. 2008. Structuration du lexique et principe d'économie : Le cas des ethniques. In Jacques Durand, Benoît Habert & Bernard Laks (eds.), *Actes du congrès mondial de linguistique française (cmlf-2008)*, 1571–1585. Paris: ILF.
- Roché, Michel. 2017. Un exemple de réseau constructionnel : ethniques, toponymes, gentilés.

- Schalchli, Gauvain & Gilles Boyé. 2018. Paradigms and syncretism in derivation: the case of ethnics in French. *Lingue e linguaggio* 2018(2). 197–216.
- Strnadová, Jana. 2014. *Les réseaux adjectivaux : sur la grammaire des adjectifs dénominaux en français*. Université Paris Diderot / Univerzita Karlova V Praze Thèse de doctorat.