

# A lexico-paradigmatic analysis of Russian demonyms

NATALIA BOBKOVA, FABIO MONTERMINI

CLLE-ERSS, CNRS & Université de Toulouse 2 Jean Jaurès

*{natalia.bobkova,fabio.montermini}@univ-tlse2.fr*

ISM<sub>o</sub> 2023, 13-15 September 2023 – Nancy

# Overview

- 1 Research object and questions
- 2 Theoretical framework
- 3 The data
- 4 Analysis
  - Paradigms
  - Exponents
- 5 Conclusion and perspectives

# Research object and questions

- Recent work in derivational morphology on demonyms in various languages:
  - form tight and regular morpholexical networks (cf. Roché [2008]; Roché [2017]; Schalchli and Boyé [2018]);
  - often make use of a large spectrum of morphological exponents → good testing ground for affixal rivalry (cf. Roché and Plénat [2016]; Thuilier et al. [2023]; Huygevelde et al. [2023]);
  - peculiar interactions between inflection and derivation, thus calling into question the frontier between the two (cf. Tuite [1995]; Schalchli and Boyé [2018]);

## Research object and questions

- As in other Slavic languages, demonymic nouns and adjectives are clearly distinct lexemes in Russian, each following a specific declension pattern;
- Adjectives consistently display the suffix *-sk* (also used for relational A); Fem. N consistently display (at least) the suffix *-k*; Masc. N display a larger variety of construction strategies:

	<u>Place N</u>	<u>Adjective</u>	<u>Masculine N</u>	<u>Feminine N</u>
(1)	Burjatij(a)	burjatsk(ij)	burjatØ	burjatk(a)
(2)	Volgograd	volgogradsk(ij)	volgogradec	volgogradk(a)
(3)	Kirov	kirovsk(ij)	kirovčanin	kirovčank(a)
(4)	Moskva	moskovsk(ij)	moskvič	moskvčk(a)
(5)	Zelenogorsk	zelenogorsk(ij)	zelenogorec	zelenogork(a)

# Research object and questions

- How to model the (semantic and formal) relations between the different words (lexemes) that compose the lexical network of Russian demonyms?
- How do inflectional and derivational morphology interact in shaping this network?
- To which extent productive morphology is influenced by the attested data? Which schemas are active / productive for speakers?

# Theoretical framework

- Paradigmatic morphology: multiple lexeme interactions within lexical networks are more relevant than local w:w relations to account for semantic and formal properties of complex lexemes (Bonami and Strnadová [2018]; Fradin [2018]; Fradin [2020] a.o.; and on ethnics: Roché [2017]; Schalchli and Boyé [2018]...)
- Constraint-based Morphology: derivational schemas are better viewed as the outcome of different, possibly conflicting, constraints (Hathout [2009]; Roché and Plénat [2014]...);

# Theoretical framework

- In particular, traditional exponents (suffixes) are themselves viewed as constraints on the formal properties of derived lexemes (cf. “templates” in Vihman [2010], Vihman and Croft [2007]);
- Such notions as “affix” and “stem” correspond to generalizations that emerge from the existing lexicon and the interaction between different constraints (vs. representations).

# The data

- Database of 4,438 demonyms from 1,412 toponyms in Russian:
  - Russian city names: 3,523 demonyms / 1,233 toponyms  
(sources: Babkin [1964];  
[https://ru.wikipedia.org/wiki/Список\\_городов\\_России](https://ru.wikipedia.org/wiki/Список_городов_России))
  - Foreign toponyms (mainly countries): 915 demonyms / 279 toponyms  
(source: Russian Wikipedia)
- Only kept demonyms that had at least one attestation in the Russian National Corpus (<https://ruscorpora.ru/>) or one attestation in discourse in Google.



# The data

- Global picture

- Adjectives straightforwardly display the suffix *-sk*:  
*volgogradsk(ij)*, *zelenogorsk(ij)*
- Feminine nouns always display (at least) the suffix *-k*:  
*volgogradk(a)*, *zelenogork(a)*
- Masculine nouns display in most cases either the suffix *-c* ([ts]) or the suffix *-anin*:  
*volgogradc(y)*<sub>M.NOM.PL.</sub>, *kirovčanin*

# The data

- Finer details

- In both the masculine and the feminine nouns the suffix has the form C when it is followed by an inflectional suffix and the form VC when it is word-final (two different allomorphic stems):

(6) volgograd**c**(y)<sub>M.NOM.PL</sub>    volgograd**k**(a)<sub>F.NOM.SG</sub>

(7) volgograd**ec**<sub>M.NOM.SG</sub>    volgograd**ok**<sub>F.GEN.PL</sub>

# The data

- Finer details

- Demonyms in *-anin* also display two allomorphic stems (SG. vs. PL.):

- (8)
- |    |   |   |
|----|---|---|
| a. | kirov <b>čanin</b> <sub>M.NOM.SG</sub>  | kirov <b>čanin(a)</b> <sub>M.GEN.SG</sub> |
| b. | kirov <b>čan(e)</b> <sub>M.NOM.PL</sub> | kirov <b>čan</b> <sub>M.GEN.PL</sub>      |

- The sequence *-anin/-an* is systematically preceded by a palatal or a soft consonant:

- (9)
- |    |          |   |                    |
|----|----------|---|--------------------|
| a. | Kirov    | → | kirov <b>čanin</b> |
| b. | Koroč(a) | → | koroč <b>čanin</b> |
| c. | Vyborg   | → | vybor <b>žanin</b> |
| d. | Kol(a)   | → | kol <b>janin</b>   |

# The data

- Additional problems:

- in some cases (as in other languages) the direction of derivation is reversed (cf. (1));
- the Fem. N is sometimes constructed directly on the toponym (*Volgograd* → *volgogradk(a)*) and sometimes on the Masc. N (*moskvič* → *moskvičk(a)*);
- some derivatives display further phonological material (“extension”) preceding the suffix:

(10)	Dn(o)	d <b>novsk</b> (ij)	d <b>nov</b> ec	d <b>novk</b> (a)
(11)	Korsik(a)	korsik <b>ansk</b> (ij)	korsik <b>an</b> ec	korsik <b>ank</b> (a)

# The data

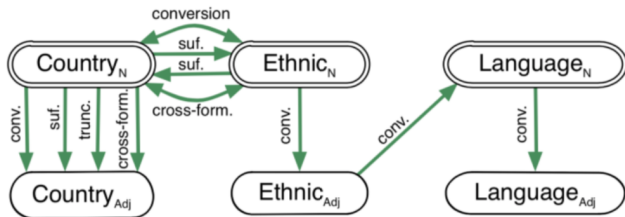
- some of the affixes used to construct demonyms are homophonous to 'affixoid' material which is encountered in toponyms, thus making some of them not 'constructed' *stricto sensu*:
  - toponyms in *-k(a)*:
 

(12)	Pallasovk(a)	pallasov <b>sk</b> (ij)	pallasove <b>c</b>	pallasov <b>k</b> (a)
(13)	Jamajk(a)	jamaj <b>sk</b> (ij)	jamaj <b>ec</b>	jamaj <b>k</b> (a)
  - toponyms in *-sk*:
 

(14)	Zelenogorsk	zelenogor <b>sk</b> (ij)	zelenogore <b>c</b>	zelenogor <b>k</b> (a)
(15)	Dzeržinsk(ij)	dzeržin <b>sk</b> (ij)	dzeržine <b>c</b>	dzeržin <b>k</b> (a)
- Note: the toponym in (14) is inflected as a (class 1) noun, that in (15) is inflected as a (masc.) adjective. Their etymological source may be diverse, including back-formation from demonyms (cf. Cexanovič [2007]), but synchronically the corresponding Adjs. all appear as (stem) conversions.

## Analysis: paradigms

- In French (and other Romance lgs.) ethnic nouns and adjectives are systematically homophonous (*italien<sub>N</sub>* / *italien<sub>A</sub>*). Schalchli and Boyé [2018] (elaborating on Roché [2008]) propose a representation aiming at accounting for the apparent polysemy of ethnic adjectives and for the derivational links between the lexemes that constitute the derivational paradigm:



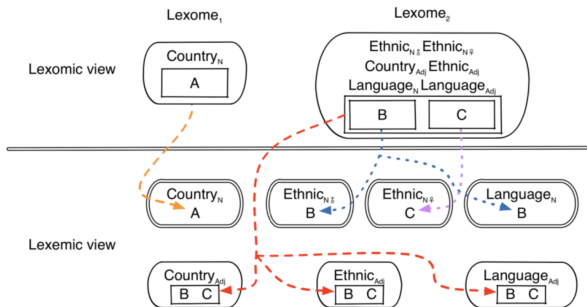
## Analysis: paradigms

- According to their analysis, French has three distinct ethnic adjectives each linked semantically to a different noun:

- (16) a. la péninsule italienne  
'the Italian peninsula' ('related to Italy')
- b. le tempérament italien  
'the Italian character' ('related to Italians')
- c. la grammaire italienne  
'the Italian grammar' ('related to the Italian language')

# Analysis: paradigms

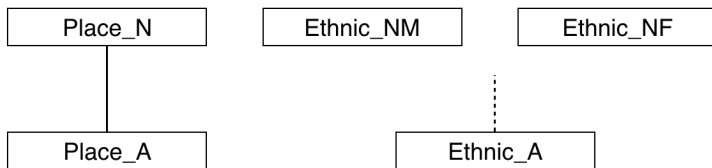
- Elaborating on the observation that masculine and feminine nouns and adjectives have systematically identical forms (what they call “derivational syncretism”), they propose to gather all these lexemes under a unique set they call “lexome”, that is instantiated in the lexicon by different lexemes:





## Analysis: paradigms

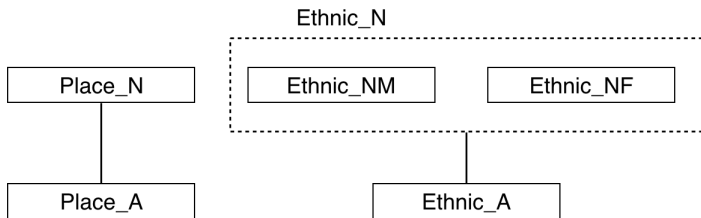
- In Russian, where the masculine and feminine noun are not homophonous, we have the problem of linking Schalchli and Boyé [2018]’s “Ethnic Adj.” to a base noun:



- Note: language nouns (and Adjs.) are left apart, since they are systematically realized through the periphrasis Adj. + *jazyk* ('language').

## Analysis: paradigms

- We can imagine to link it to a unit which is undefined for gender and which is superordinate to the two actual lexemes; however it is not clear what the nature of this 'meta-lexeme' would be:



## Analysis: paradigms

- Rather, we consider that the Russian data call into question the distinction between three discrete (systematically homophonous) adjectival lexemes based on their meaning, and suggests that ethnic adjectives are better analyzed as flexible items that can express a large spectrum of meanings, including the three in question (multiple semantic connexions):

- (17) a. les restaurants italiens à Paris  
'the Italian restaurants in Paris'  
(‘related to the Italian kitchen’ (neither))
- b. l’essai fidjien a été refusé (Web)  
'the Fijian try was refused'  
(‘by the team of Fiji’ or ‘by the Fijians’ or... (>1))

# Analysis: exponents

- Whereas adjectives are straightforwardly formed, nouns pose various issues for the analysis:
  - Rivalry in masculine noun formation: how is the affix chosen?
  - Linking: which is the base of feminine nouns?

(18)	a.	Volgograd	volgograd <b>c</b> (y)	volgograd <b>k</b> (a)
	b.	Kirov	kirov <b>č</b> an(e)	kirov <b>č</b> an <b>k</b> (a)
	c.	Moskva	moskvi <b>č</b>	moskvi <b>č</b> k(a)

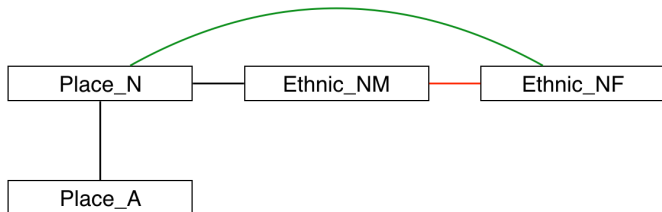
# Analysis: exponents

- Cases like (18b-c) seem to suggest a derivation from the masculine noun, while cases like (18a) seem to suggest a direct derivation from the toponym:

(18)	a.	Volgograd	volgograd <b>c</b> (y)	volgograd <b>k</b> (a)
	b.	Kirov	kirovč <b>an</b> (e)	kirovč <b>an</b> k(a)
	c.	Moskva	moskvič	moskvič <b>k</b> (a)

## Analysis: exponents

- Cases like (18b-c) seem to suggest a derivation from the masculine noun (red), while cases like (18a) seem to suggest a direct derivation from the toponym (green):



(A third hypothesis, compatible with the red path, would be that the two suffixes *-c* and *-k* are in a relation of mutual deletion → independent evidence?).

## Analysis: exponents

- Arguments in favor of the hypothesis of a direct, parallel, derivation of masc. and fem. nouns from the toponym:
  - The *-c/-k* pattern is available for other types of bases (often (pseudo-)proper names, including neologisms (cf. (20.b-c), Zavarzina [2017]):

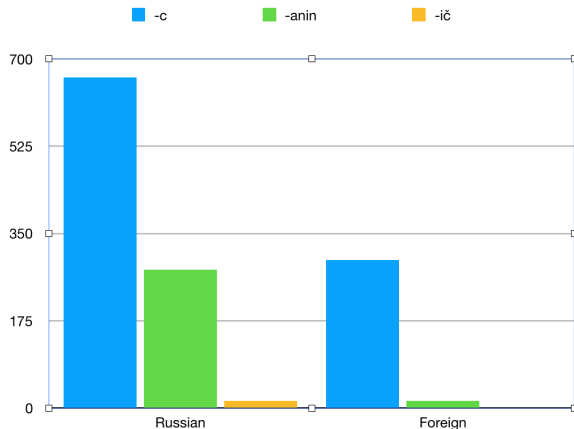
(19)	a.	Komsomol 'communist youth'	→	komsomolec	komsomolk(a)
	b.	pensionnyj fond 'pension fund'	→	fondovec	fondovk(a)
	c.	Omon 'special police'	→	omonovec	omonovk(a)

# Analysis: exponents

- Arguments in favor of the hypothesis of a direct, parallel, derivation of masc. and fem. nouns from the toponym:
  - Overwhelming majority of the  $-c/-k$  pattern in demonyms from foreign toponyms;
  - likely to be more recent / less lexicalized, i.e. better representatives of the productive use of derivational schemas by speakers in synchrony.



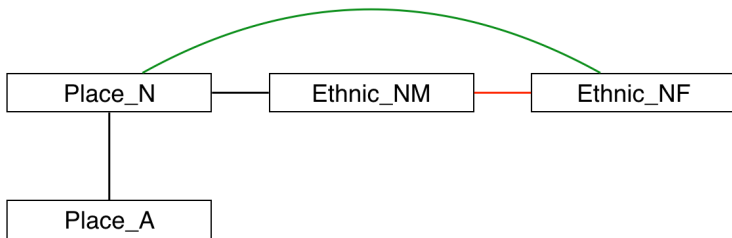
# Analysis: exponents



Proportion of the main patterns for M and F Ns in the two sub-databases.

# Analysis: exponents

- Arguments in favor of the hypothesis of a direct, parallel, derivation of masc. and fem. nouns from the toponym:



# Conclusion and perspectives

- First observations of a database of corpus-extracted ethnic nouns and adjectives;
- Assess the validity of derivational paradigms identified for other, structurally different, languages;
- Determine which schemas are available for speakers in synchrony.

# Conclusion and perspectives

- More accurate statistical analyses on the properties of bases that intervene in the choice of the affix (rivalry in ethnic masculine nouns);
- Measure the strength of paradigmatic connexions for production / interpretation;
- Experimental assessment of the speakers' intuitions both in production and in interpretation;
- More thorough analysis of variability, e.g. in connection with frequency.

# References I

- Aleksandr M. Babkin. *Словарь названий жителей (РСФСР)*. Sovetskaja Ėnciklopedija, 1964.
- Olivier Bonami and Jana Strnadová. Paradigm structure and predictability in derivational morphology. *Morphology*, 29(2):167–197, 2018.
- Marianna A. Сеханови́ч. Суффикс -ск- в русской топонимии и проблема наложения морфем в оттопонимических прилагательных. *Вестник ПСТГУ*, 9(3), 2007.
- Bernard Fradin. Paradigms and the role of series in derivational morphology. *Lingue e linguaggio*, XVII(2), 2018.
- Bernard Fradin. Characterizing derivational paradigms. In Alexandra Bagasheva, Jesús Fernández-Domínguez, and Cristina Lara-Clares, editors, *Paradigmatic Relations in Derivational Morphology*, pages 49–84. Brill, 2020.
- Nabil Hathout. Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie, 2009.
- Marie Huygevelde, Ridvan Kayirici, Olivier Bonami, and Barbara Hemforth. Affix rivalry in french demonyms: an experimental approach. In *This conference*, 2023.
- Michel Roché. Structuration du lexique et principe d'économie : le cas des ethniques. In Durand, Jacques, Habert, Benoît, and Laks, Bernard, editors, *Actes du Congrès Mondial de Linguistique Française.*, pages 1571–1585. Institut de Linguistique Française, 2008.
- Michel Roché. Un exemple de reseau constructionnel : ethniques, toponymes, gentiles. *Manuscript*, 2017.
- Michel Roché and Marc Plénat. Le jeu des contraintes dans la sélection du thème pré-suffixal. In Neveu, Franck, Blumenthal, Peter, Hriba, Linda, Gerstenberger, Annette, Meinschäfer, Judith, and Prévost, Sophie, editors, *Actes du 4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 juillet 2014*, pages 1863–1878. Institut de Linguistique Française, 2014.
- Michel Roché and Marc Plénat. De l'harmonie dans la construction des mots français. In Neveu, Franck, Bergounioux, Gabriel, Côté, Marie-Hélène, Fournier, Jean-Marie, Hriba, Linda, and Prévost, Sophie, editors, *Actes du 5e Congrès Mondial de Linguistique Française*. Institut de Linguistique Française, 2016.
- Gauvain Schalchli and Gilles Boyé. Paradigms and syncretism in derivation: The case of ethnics in French. *Lingue e linguaggio*, XVII(2):197–215, 2018.

# References II

- Andrea D. Sims. Slavic morphology: Recent approaches to classic problems, illustrated with Russian. *Journal of Slavic Linguistics*, 25(2):489–542, 2017.
- Juliette Thuilier, Delphine Tribout, and Marine Wauquier. Affix rivalry in french demonym formation: The role of linguistic and non-linguistic parameters. *Word Structure*, 16(1), 2023.
- Kevin Tuite. The declension of ethnonyms in english. In *Proceedings of the Twenty-First Annual Meeting of the Berkeley Linguistics Society*. Linguistic Society of America, 1995.
- Marilyn Vihman. Phonological templates in early words. In Fougeron, Cécile, D'Imperio, Mariapaola, and Vallée, Nathalie, editors, *Laboratory Phonology 10. Tenth Conference in Laboratory Phonology (LabPhon 10), 29 Jun - 01 Jul 2006*, pages 261–284. Mouton de Gruyter, 2010.
- Marilyn Vihman and William Croft. Phonological development: toward a "radical" templatic phonology. 45(4):683–725, 2007.
- Galina A. Zavarzina. Словообразовательные общественно-политические неологизмы в русском языке новейшего периода. *Русистика*, (2):5–10, 2017.