

# **Morphological epenthesis in Serbo-Croatian noun inflection**

**Andrija Petrovic**

SBU/RU MLRG, Mar 2 2021

# Introduction

- Class I:

zavod	zavod-a
institute.NOM.SG.M	institute.GEN.SG.M
sel-o	sel-a
village.NOM.SG.N	village.GEN.SG.N
tele	telet-a
calf.NOM.SG.N	calf.GEN.SG.N
- Class II:

žen-a	žen-e
woman.NOM.SG.F	woman.GEN.SG.F
- Class III:

rečf	rečf-i
word.NOM.SG.F	word.GEN.SG.F

# Introduction

- Class I:

zavod	zavod-a
institute.NOM.SG.M	institute.GEN.SG.M
sel-o	sel-a
village.NOM.SG.N	village.GEN.SG.N
tele	telet-a
calf.NOM.SG.N	calf.GEN.SG.N
- Class II:

žen-a	žen-e
woman.NOM.SG.F	woman.GEN.SG.F
- Class III:

reč	reč-i
word.NOM.SG.F	word.GEN.SG.F

# Introduction

- Class I:

zavod	zavod-a
institute.NOM.SG.M	institute.GEN.SG.M
sel-o	sel-a
village.NOM.SG.N	village.GEN.SG.N
tele	telet-a
calf.NOM.SG.N	calf.GEN.SG.N
- Class II:

žen-a	žen-e
woman.NOM.SG.F	woman.GEN.SG.F
- Class III:

reč	reč-i
word.NOM.SG.F	word.GEN.SG.F

morphologically  
conditioned consonant  
insertion  
—  
repairing vowel-final  
stems

# Introduction

- part of a wider algorithmic process of noun inflection
  - Boolean Monadic Recursive Schemes (BMRSs)
  - analyzed as logical transductions on strings
    - stem + [Case Num]  $\rightarrow$  inflected word       $\sim$  realization rules
  - hierarchical ordering of more specific and less specific blocking and licensing structures
    - $\sim$  realization rules intrinsically ordered by Pāṇini's principle (cf. Stump 2001)
- possibility to compose BMRSs into a complete inflectional system
  - algorithm of inflectional class assignment

# Morphologically conditioned consonant insertion

- SC stems are normally C-final; those that are not are repaired via consonant insertion  
→ if another suffix follows immediately

otherwise the word is vowel-final (bare stem) and considered well-formed (as the nom.sg. suffix is *-o/-e*)

	tele	‘calf.NOM.SG’
	telet-a	‘calf.GEN.SG.’
/tele + ji/	[teletɕi]	‘calf-like(ADJ.)’

# Morphologically conditioned consonant insertion

- *t* is not the general epenthetic consonant in this language

/zaova/            [zaova]            ‘sister-in-law’

/violina/            [vijolina]            ‘violin’

- *t* epenthesized only to repair vowel-final stems, does not depend on vowel hiatus

/tele + ji/        [teletɕi] ‘calf-like’ vs.        /koz + ji/        [kozji] ‘goat-like’

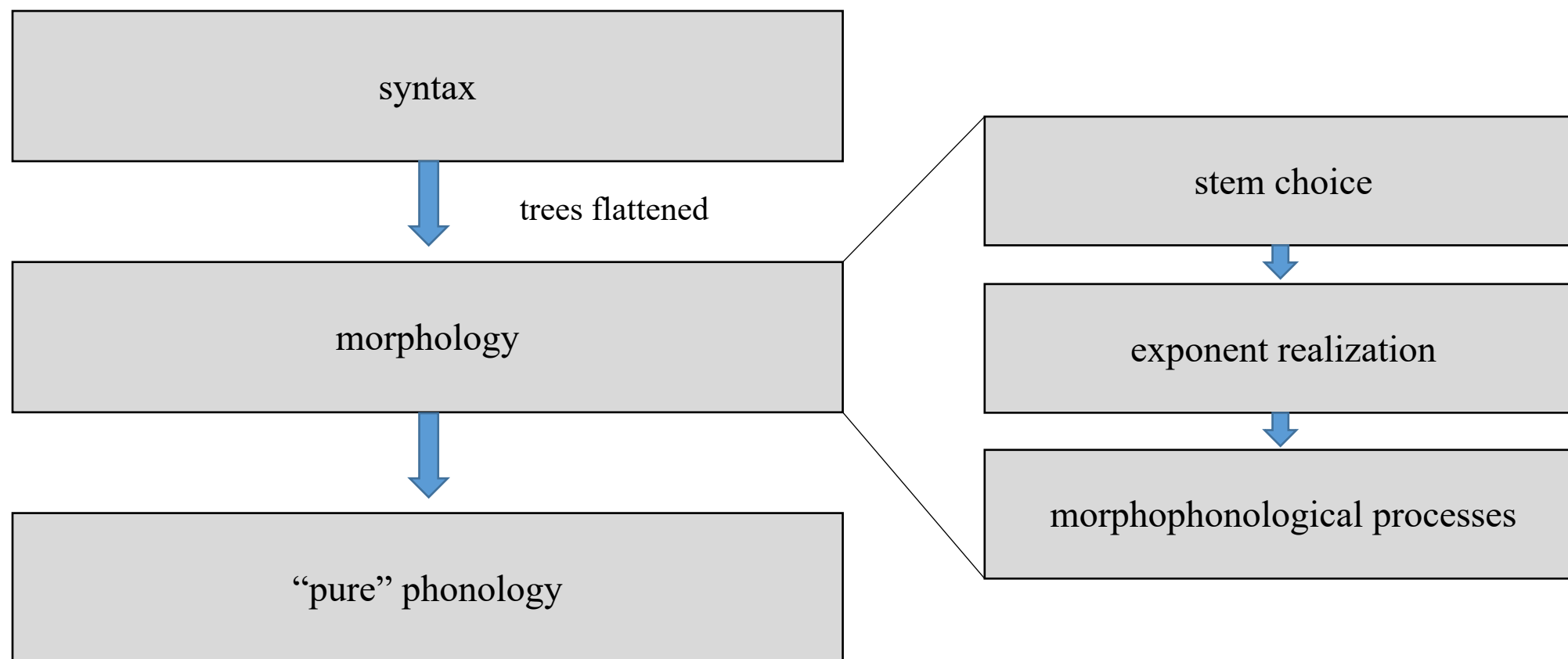
/sirtɕe + ni/    [sirtɕetni] ‘acetic’ vs.        /plod + ni/    [plodni] ‘fertile’

# The nature of morphological processes

- morphological processes are regular (Karttunen et al. 1992), most often subregular (Chandlee 2017)
- DM is a tree-based theory, but can be formalized over strings (Ermolaeva & Edmiston 2018)
  - yield function flattens trees, retaining necessary boundary symbols
- In terms of computational power, W&P morphology is equivalent to a collection of regular relations (Karttunen 2003)



# Proposed model



# Analysis

- *Boolean Monadic Recursive Schemes* (BMRS)
  - incorporates the observation that phonology is regular (Johnson 1972; Kaplan & Kay 1994; Heinz 2018)
  - morphology, too, is regular (Karttunen et al. 1992), most often subregular (Chandlee 2017)
  - unlike FSTs, captures linguistically significant generalizations – we can use representations that are common in linguistics
    - phonological, morphosyntactic features
    - Pāṇini's Principle (= Elsewhere condition) has no obvious finite-state implementation (Karttunen 2003), whereas it is directly captured with BMRS's 'if...then...else' syntax

# Analysis

- primitives:
  - *boolean values*  $\top$  and  $\perp$
  - *monadic predicates*  $P(t)$  – take a *single* argument  $t$ , and return  $\top$  or  $\perp$

$$\mathbb{V} = \{a, e, i, o, u\}$$

$$\mathbb{C} = \{b, \mathfrak{b}, \mathfrak{t}, \mathfrak{d}, \mathfrak{g}, \mathfrak{z}, f, g, h, j, k, l, \wedge, m, n, \mathfrak{n}, p, r, s, \int, t, v, z, 3\}$$

$$\mathbb{M} = \{[nom], [gen], [dat], [acc], [voc], [ins], [loc], [sg], [pl]\}$$

$$\Sigma = \mathbb{V} \cup \mathbb{C} \cup \mathbb{M} \qquad \Sigma_{\times} = \Sigma \cup \{\times, \times, +\}$$

# Analysis

- primitives:
  - *boolean values*  $\top$  and  $\perp$
  - *monadic predicates*  $P(t)$  – take a *single* argument  $t$ , and return  $\top$  or  $\perp$

$$\mathbb{V} = \{a, e, i, o, u\}$$

$$\mathbb{C} = \{b, t, f, d, g, h, j, k, l, \wedge, m, n, p, r, s, \int, v, z, 3\}$$

$$\mathbb{M} = \{[nom], [gen], [dat], [acc], [voc], [ins], [loc], [sg], [pl]\}$$

$$\Sigma = \mathbb{V} \cup \mathbb{C} \cup \mathbb{M} \quad \Sigma_{\bowtie} = \Sigma \cup \{\bowtie, \ltimes, +\}$$

$$\mathcal{I} = \{a(t), \dots, 3(t), [nom](t), \dots, [pl](t), \bowtie(t), \ltimes(t), +(t)\}$$

$$\mathcal{O} = \{a_o(t), \dots, 3_o(t), [nom]_o(t), \dots, [pl]_o(t), \bowtie_o(t), \ltimes_o(t), +_o(t)\}$$

# Analysis

- no right word boundary symbol  $\rightarrow$  stems are (contiguous) strings of characters occupying indices belonging to a *countably infinite* set of indices

$$\mathbf{S} = \langle D; \sigma_1, \sigma_2, \dots, p, s \rangle$$

$$D = \{1, 2, \dots\}$$

- the right word boundary is assigned at the end of the word only after the inflectional exponents are realized (look out for  $\bowtie(\mathbf{x})$  in the following slides!)
- (I do this to avoid using copy sets, which might not be a great idea.)

# Analysis

- In BMRS, we can refer to both input and output predicates, and define new ones

<u>input</u>							
1	2	3	4	5	6	7	...
×	t	e	l	e	+	[dat sg]	...
<u>output</u>							
t e l e t u							

- Examples of defining new input predicates: case/number feature combinations, syncretism, defining the class of FCs

# Analysis

- necessary combinations of case and number features available for neuter nouns:

$[\text{nom/acc/voc sg}](x) = \text{if } [\text{nom/acc/voc}](x) \text{ then } [\text{sg}](x) \text{ else } \perp$   
 $[\text{dat/loc sg}](x) = \text{if } [\text{dat/loc}](x) \text{ then } [\text{sg}](x) \text{ else } \perp$   
 $[\text{ins sg}](x) = \text{if } [\text{ins}](x) \text{ then } [\text{sg}](x) \text{ else } \perp$   
 $[\text{nom/acc/voc pl}](x) = \text{if } [\text{nom/acc/voc}](x) \text{ then } [\text{pl}](x) \text{ else } \perp$   
 $[\text{gen pl}](x) = \text{if } [\text{gen}](x) \text{ then } [\text{pl}](x) \text{ else } \perp$   
 $[\text{dat/loc pl}](x) = \text{if } [\text{dat/loc}](x) \text{ then } [\text{pl}](x) \text{ else } \perp$   
 $[\text{ins pl}](x) = \text{if } [\text{ins}](x) \text{ then } [\text{pl}](x) \text{ else } \perp$

(one would expect  $14 - 7 \text{ cases} \times 2 \text{ numbers}$  – but we do not have 14 distinct phonological forms of overt suffixes)

# Analysis

- (non-directional) syncretism

$[\text{dat/loc}](x) = \text{if } [\text{dat}](x) \text{ then } \tau \text{ else } [\text{loc}](x)$

$[\text{nom/acc/voc}](x) = \text{if } [\text{nom}](x) \text{ then } \tau \text{ else}$   
 $\text{if } [\text{acc}](x) \text{ then } \tau \text{ else } [\text{voc}]$

$[\text{gen/ nom/acc/voc pl}](x) = \text{if } [\text{gen}](x) \text{ then } \tau$   
 $\text{else } [\text{nom/acc/voc pl}](x)$

$[\text{dat/loc pl/ins pl}](x) = \text{if } [\text{dat/loc pl}](x) \text{ then } \tau$   
 $\text{else } [\text{ins pl}](x)$



# Analysis

- effect of fronting consonants (i.e. fronting of  $\text{ɔ}$  to  $\text{e}$  after stem-final FCs)

(FC = C[cor, -ant] +  $\text{ʌ}$ )

$[\text{FC}](x) =$

- if  $\text{ʌ}(x)$  then  $\text{ʌ}$  else
- if  $\text{ɹ}(x)$  then  $\text{ʌ}$  else
- if  $\text{tʌ}(x)$  then  $\text{ʌ}$  else
- if  $\text{dʒ}(x)$  then  $\text{ʌ}$  else
- if  $\text{dʒ}(x)$  then  $\text{ʌ}$  else
- if  $\text{j}(x)$  then  $\text{ʌ}$  else
- if  $\text{ʌ}(x)$  then  $\text{ʌ}$  else
- if  $\text{j}(x)$  then  $\text{ʌ}$  else
- if  $\text{j}(x)$  then  $\text{ʌ}$  else  $\text{ʌ}(x)$

# Analysis

- $t$ -insertion

$\text{outseg}(x) =$  if  $a_o(x)$  then  $\top$  else  
          if  $b_o(x)$  then  $\top$  else  
          if  $c_o(x)$  then  $\top$  else  
          (...)  
          if  $z_o(x)$  then  $\top$  else  $z_o(x)$

$t_o(x) =$  if  $+(x)$  then  
          if  $e(p(x))$  then  $\text{outseg}(s(x))$  else  $\perp$   
          else  $t(x)$

# Analysis

- Output predicates: ( $\sim$  realization rules)

$u_o(x) = \text{if } [\text{dat/loc sg}](x) \text{ then } \top \text{ else } u(x)$   
 $a_o(x) = \text{if } [\text{gen/ nom/acc/voc pl}](x) \text{ then } \top \text{ else}$   
           $\text{if } [\text{gen pl}](p(x)) \text{ then } \top \text{ else}$   
           $\text{if } [\text{ins pl}](p(p(x))) \text{ then } \top \text{ else } a(x)$   
 $o_o(x) = \text{if } [\text{nom/acc/voc sg}](x) \text{ then } \top \text{ else}$   
           $\text{if } [\text{ins sg}](x) \text{ then } \top \text{ else } o(x)$   
 $m_o(x) = \text{if } [\text{ins}](p(x)) \text{ then } \top \text{ else } m(x)$   
 $i_o(x) = \text{if } [\text{ins pl}](x) \text{ then } \top \text{ else } i(x)$   
 $e_o(x) = \text{if } o_o(x) \text{ then}$   
           $\text{if } +(p(x)) \text{ then } [\text{pal}](p(p(x))) \text{ else } \perp$   
           $\text{else } e(x)$

$l_o(x) = l(x)$

$b_o(x) = b(x)$

(etc., for  $\{\text{t}, \text{t}^h, \text{t}^s, \text{d}, \text{d}^h, \text{d}^s, \text{f}, \text{g}, \text{h}, \text{j}, \text{k}, \text{l}, \text{ʎ, n}, \text{n}, \text{p}, \text{r}, \text{s}, \text{ʃ}, \text{v}, \text{z}, \text{z}^h\}$ )

# Analysis

<u>input</u>								
1	2	3	4	5	6	7	8	9
×	t	e	l	e	+	[ins sg]		...
<u>output</u>								
	t	e	l	e	t	o	m	

- Output predicates:

$u_o(x) = \text{if } [\text{dat/loc sg}](x) \text{ then } \top \text{ else } u(x)$

$a_o(x) = \text{if } [\text{gen/ nom/acc/voc pl}](x) \text{ then } \top \text{ else}$   
            $\text{if } [\text{gen pl}](p(x)) \text{ then } \top \text{ else}$   
            $\text{if } [\text{ins pl}](p(p(x))) \text{ then } \top \text{ else } a(x)$

$o_o(x) = \text{if } [\text{nom/acc/voc sg}](x) \text{ then } \top \text{ else}$   
            $\text{if } [\text{ins sg}](x) \text{ then } \top \text{ else } o(x)$

$m_o(x) = \text{if } [\text{ins}](p(x)) \text{ then } \top \text{ else } m(x)$

$i_o(x) = \text{if } [\text{ins pl}](x) \text{ then } \top \text{ else } i(x)$

$e_o(x) = \text{if } o_o(x) \text{ then}$   
            $\text{if } +(p(x)) \text{ then } [\text{pal}](p(p(x))) \text{ else } \perp$   
            $\text{else } e(x)$

$l_o(x) = l(x)$

$b_o(x) = b(x)$

(etc., for {t, t̃, t̂, d, d̃, d̂, f, g, h, j, k, l, l̃, n, ñ, p, r, s, s̃, v, z, z̃})

# Analysis

$\bowtie(x) =$  if [nom sg](x) then  $e(p(x))$  else  
if [nom/acc/voc sg]( $p(x)$ ) then  $\top$  else  
if [dat/loc sg]( $p(x)$ ) then  $\top$  else  
if [gen/ nom/acc/voc pl](x) then  $\top$  else  
if [ins sg]( $p(p(x))$ ) then  $\top$  else  
if [dat/loc pl/ins pl]( $p(p(p(x)))$ ) then  $\top$  else  $\perp$

$\text{out}(x) =$  if  $+(x)$  then  $t_o(x)$  else  
if  $\bowtie(x)$  then  $\perp$  else  
if  $\bowtie(x)$  then  $\perp$  else  $\top$

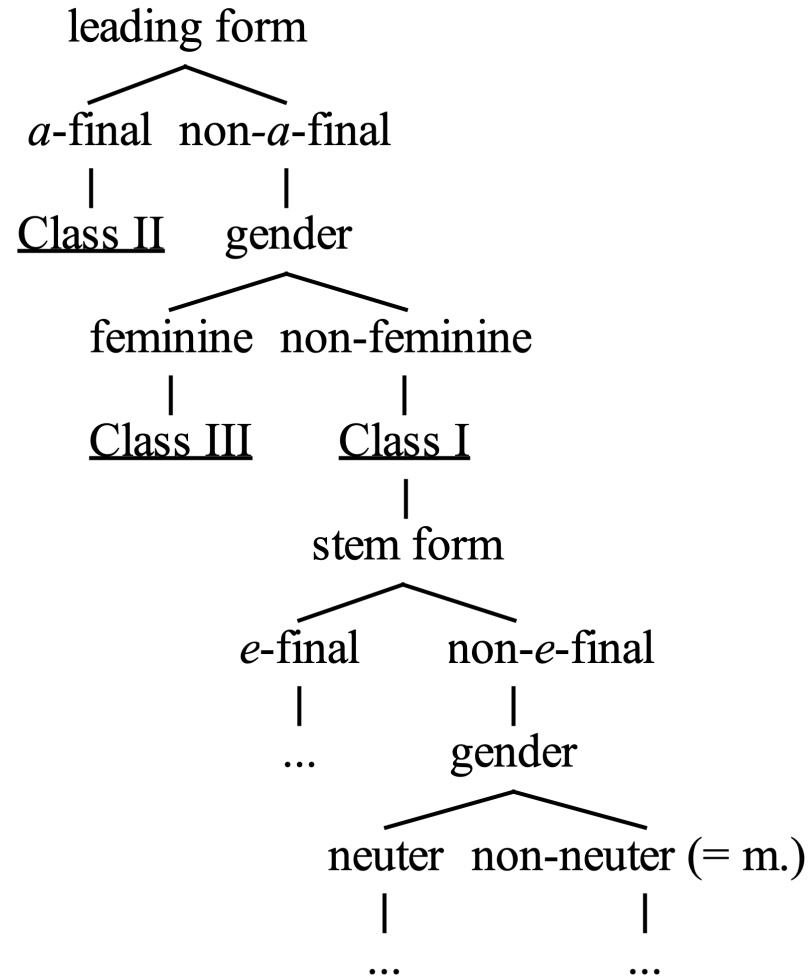
# Analysis

- no copy sets = no epenthesis/insertion of material longer than 1 symbol ☹
- is there a smart way to do this (+ more → non-concatenative morphology) without copy sets? The non-computationally minded often find them counterintuitive

# Class membership assignment in SC

- Class membership assignment in SC can be modeled as a predictable process if we take into account a *leading form/base* (Wurzel 1990, McCarthy 2005, Albright 2008)
- This I argue to be the nominative singular: it is the most frequent form (subject case), and it ranks highest in the Case Hierarchy (Blake 2001).
- loanwords do not (necessarily) have an oblique form – they are incorporated into the declension system on the basis of their input form (which is borrowed as nominative singular)
- so, lexical entry = {stem, leading form, morphosyntactic info}, but the algorithm should work with a deficient lexical entry too

# Class membership assignment in SC





# Conclusions

- Inserting *t* repairs aberrant, V-final stems (whenever an overt suffix is attached)
- This happens regardless of whether that suffix is vowel- or consonant-initial
- A means to avoid stipulating listed stems – assuming unpredictable stem allomorphs would basically reduce the phenomenon to an accident
- Assuming morphological epenthesis enables us to express the generalizations explicitly and overtly

# Conclusions

- BMRSs can directly capture morphological and phonological generalizations, retaining the computationally restrictive nature of such processes
- Intuitive, easily implementable, extendable to a wider range of phenomena
- Future work: noun class membership assignment base on ‘if...then...else’ syntax; class membership does not have to be listed information

# References

- Albright, A. (2008). Inflectional Paradigms Have Bases Too. Arguments from Yiddish. In: A. Bachrach & A. Nevins (Eds.), *The Bases of Inflectional Identity*. Oxford: Oxford University Press.
- Bhaskar, S., Chandlee, J., Jardine, A., & Oakden, C. (2020). Boolean monadic recursive schemes as a logical characterization of the subsequential functions. In *International Conference on Language and Automata Theory and Applications* (pp. 157-169). Springer, Cham.
- Blake, B. J. (2001). *Case*. Cambridge: Cambridge University Press.
- Chandlee, J. (2017). Computational locality in morphological maps. *Morphology*, 27:599– 641.
- Chandlee, J., & Jardine, A. (2020). Recursive Schemes for Phonological Analysis. Unpublished MS. Retrieved from <http://adamjardine.net/files/chandleejardineBMRsms.pdf>
- Ermolaeva, M., & Edmiston, D. (2018). Distributed morphology as a regular relation. *Proceedings of the Society for Computation in Linguistics*, 1(1), 178-181.
- Heinz, J. (2018). The computational nature of phonological generalizations. In Hyman, L. and Plank, F., editors, *Phonological Typology*, Phonetics and Phonology, chapter 5, pages 126–195. De Gruyter Mouton.
- Johnson, C. D. (1972). *Formal aspects of phonological description*. Mouton.
- Kaplan, R. and Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*, 20:331–78.
- Karttunen, L. (2003). Computing with realizational morphology. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 203-214). Springer, Berlin, Heidelberg.
- Karttunen, L., Kaplan, R. M., & Zaenen, A. (1992). Two-level morphology with composition. In *COLING 1992 Volume 1: The 15th International Conference on Computational Linguistics*.
- McCarthy, J. (2005). Optimal Paradigms. In: L. Downing, T. Hall & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press, pp. 170–210.
- Moradi, S. (2017). *Non-canonical Epenthesis: Epenthetic quality and the role of morphonology* (ms). Stony Brook University.
- Repetti, L., Moradi, S. & Aronoff, M. (2018). *Epenthesis and Morphology in Romance*. Presented at LSRL 48, York University, Toronto.
- Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.
- Wurzel, W. U. (1990). The Mechanism of Inflection: Lexical Representation, Rules, and Irregularities. In: W. U. Dressler (Ed.), *Contemporary Morphology*. Berlin: Mouton de Gruyter.