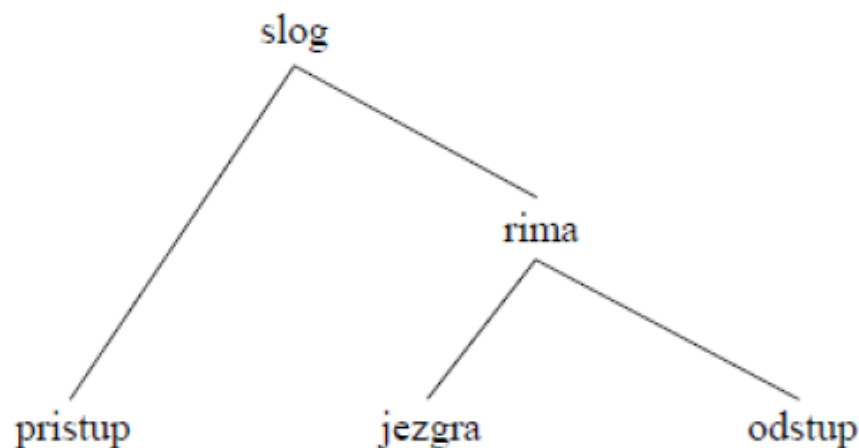


R4

Sažetak hr/eng

Prezentacija:
20 kliznica

Filmić:
65 minuta



Hrvatski mrežni subčestični (strukturni) rječnik

Josip Užarević, Suzana Molčanov, Antonio Jovanović
juzarevic@ffzg.hr; suzana.molcanov@gmail.com;
ajovanov@mathos.hr

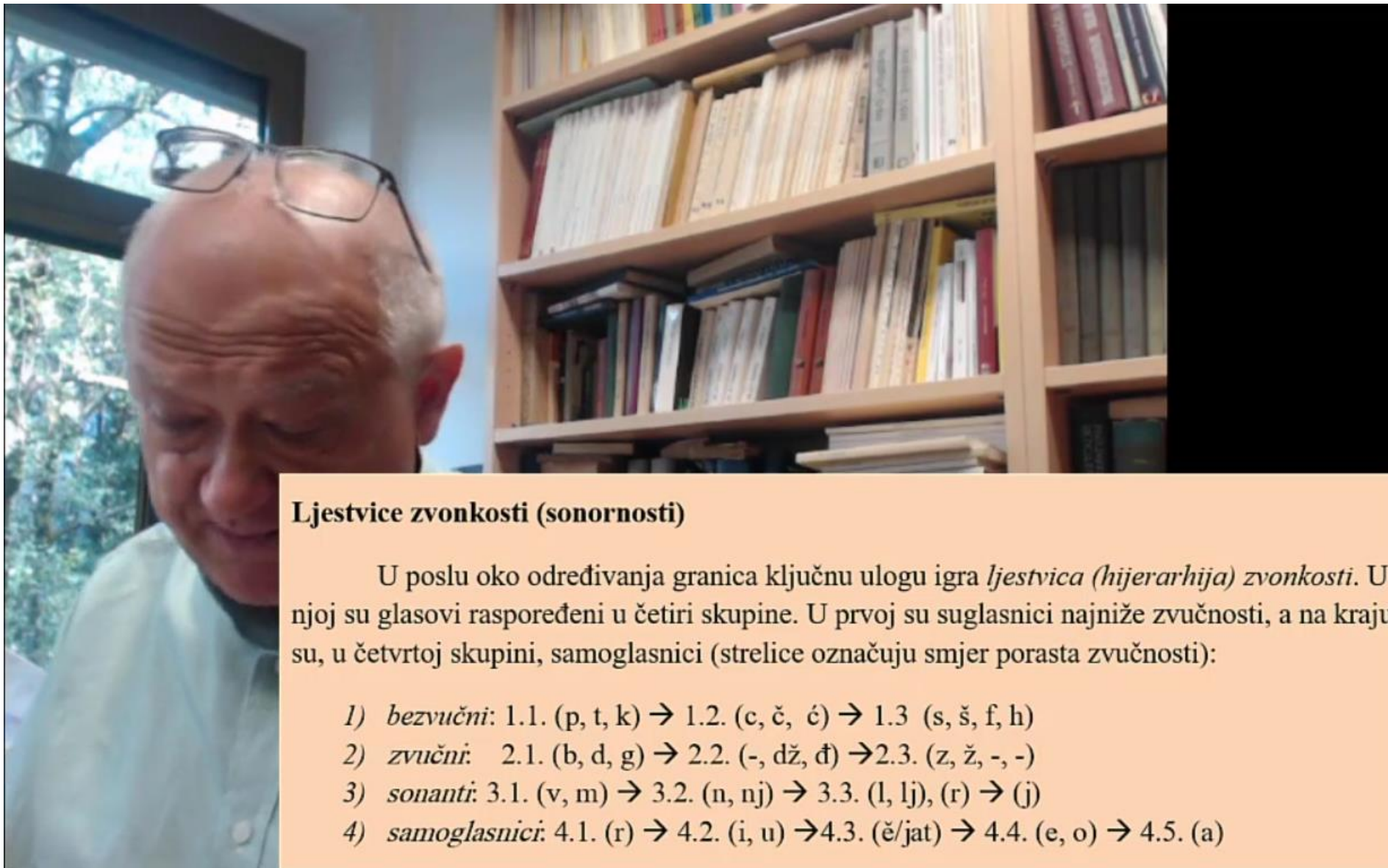
CLGF-2020

Osijek, 25.-26. rujna 2020.



❖ O slogovima

❖ O slogovniku
hrvatskoga
jezika



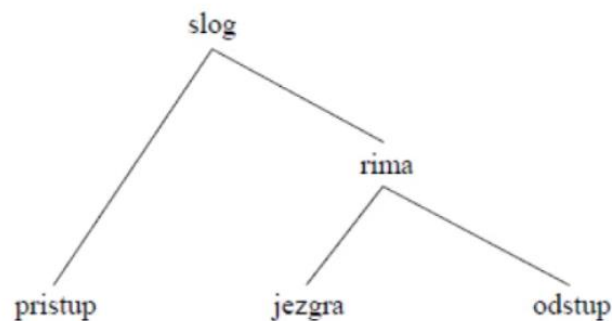
Ljestvice zvonкости (sonornosti)

U poslu oko određivanja granica ključnu ulogu igra *ljestvica (hijerarhija) zvonкости*. U njoj su glasovi raspoređeni u četiri skupine. U prvoj su suglasnici najniže zvučnosti, a na kraju su, u četvrtoj skupini, samoglasnici (strelice označuju smjer porasta zvučnosti):

- 1) *bezvučni*: 1.1. (p, t, k) → 1.2. (c, č, ć) → 1.3. (s, š, f, h)
- 2) *zvučni*: 2.1. (b, d, g) → 2.2. (-, dž, đ) → 2.3. (z, ž, -, -)
- 3) *sonanti*: 3.1. (v, m) → 3.2. (n, nj) → 3.3. (l, lj), (r) → (j)
- 4) *samoglasnici*: 4.1. (r) → 4.2. (i, u) → 4.3. (ě/jat) → 4.4. (e, o) → 4.5. (a)

- ❖ Problemi slogovnih granica
- ❖ Ljestvice zvučnosti (4 do 15 razina)

Model sloga



- Osnovni model sloga sliči apstraktnome stablu u koje se svi slogovi trebaju

Nutarnja struktura riječi (Mario Essert) Subčestice riječi

Model silabomorfema⁴

Infinitivni dočetc (silabomorfemi) tipa *-včvč(v)*
i prezentski dočetc (silabomorfemi) *-včvč(v)*

(Užarević 2013: 229-231)

Dočetc na -a		Dočetc na -e		Dočetc na -i		Dočetc na -o		Dočetc na -r		Dočetc na -u	
Inf.	Prez.	Inf.	Prez.	Inf.	Prez.	Inf.	Prez.	Inf.	Prez.	Inf.	Prez.
-ābāt > -ābān		-ēbāt > -ēbān		-ībāt > -ībān		-ōbāt > -ōbān		-ībāt > -ībān		-ūbāt > -ūbān	
-ābīt > -ābīn				-ībīt > -ībīn		-ōbīt > -ōbīn		-ībīt > -ībīn		-ūbīt > -ūbīn	
-ābljāt > -ābljān											
-ābrīt > -ābrīn											
-ācāt > -ācān		-ēcāt > -ēcān		-īcāt > -īcān		-ōcāt > -ōcān		-īcāt > -īcān		-ūcāt > -ūcān	
-ācīt > -ācīn						-ōcīt > -ōcīn				-ūcīt > -ūcīn	
-āckāt > -āckān										-ūckāt > -ūckān	
								-īcnīt > -īcnīn		-ūcnīt > -ūcnīn	
-āčāt > -āčān				-īčāt > -īčān				-īčāt > -īčān		-ūčāt > -ūčān	
-āčīt > -āčīn		-ēčīt > -ēčīn		-īčīt > -īčīn				-īčīt > -īčīn		-ūčīt > -ūčīn	
-āćāt > -āćān				-īćāt > -īćān				-īćāt > -īćān		-ūćāt > -ūćān	
				-īćkāt > -īćkān				-īćkāt > -īćkān		-ūćkāt > -ūćkān	
				-īcnīt > -īcnīn				-īcnīt > -īcnīn		-ūcnīt > -ūcnīn	
-ādāt > -ādān		-ēdāt > -ēdān		-īdāt > -īdān		-ōdāt > -ōdān		-īdāt > -īdān			
		> -ēdēn									
-ādīt > -ādīn		-ēdīt > -ēdīn		-īdīt > -īdīn		-ōdīt > -ōdīn		-īdīt > -īdīn		-ūdīt > -ūdīn	
-ādāt > -ādān		-ēdāt > -ēdān		-īdāt > -īdān							

⁴Užarević, Josip, Skup u Osijeku: "Hrvatski jezik u računalnome jezikoslovlju", 15. veljače 2019.

Model morfema³

P0-3 K2 I K1 S0-3 Sg Oblik

gdje su:

- P0-3 - prefiksi od 0 do 3 morfa
- K2 i K1 - korijeni riječi
- I - interfiks, sveza dvaju korijenaa
- S0-3 - sufiksi od 0 do 3 morfa, a Sg - gramatički sufiks

Primjeri:

P0-3	K2	I	K1	S0-3	Sg	Oblik
-	-	-	ruk	n	i	ručni
za	-	-	ruk	i	ti	zaručiti
prep-po	-	-	ruk	i	ti	preporučiti
-	ljev	o	ruk	-	-	ljevoruk
-	ljev	o	ruk	-	im	ljevorukim
-	ruk	o	tvor	in	a	rukotvorina

³Marković, Ivan (2013) *Hrvatska morfonologija*. Zagreb: Disput.

Užarević, Molčanov, Jovanović (FFZG,MathOS)

Hrvatski mrežni subčestični rječnik

Osijek, 25.-26. rujna 2020.

4 /

- ❖ Tri modela subčestičnih rječnika unutar SSF-a
- ❖ Opća slogotvorna pravila

Slogovi

rijeci.txt:

```
škripanje, škropljenje, štrecanje, ždrebešće, ždrijebe,  
ždrijelo, smrekovina, stjuardesa, škrabica, škrapa,  
škrinja, posestrimstvo, dvije, higijena, znanstveni, sjevernjak, zlopatnik,  
predodžba, usmeni, sunašće, istarski, pašće, primjedba, jezerce,  
mladićstvo, bapski, postkoncilski, mlijeko, renesansni, sportski,
```

padine.py:

```
## Padine pristupa i odstupa ###  
# '\u0247' = r (slogotvorno), '\u01cc' = nj, '\u01c9' = lj, '\u01c6' = dž,  
\u02c2' = ije, '\u02c5' = je  
pristupi = [  
[
```

☐ Ispis međurezultata ☒ Proširenje

Obradi »

Reset

tezine.py:

```
### TABLICA ZVONAČNOSTI ###  
  
tezine = {  
    'p': 1, 't': 1, 'k': 1, 'c': 1, 'č': 1, 'ć': 1,  
    'b': 2, 'd': 2, 'g': 2, 'đ': 2, '\u01c6': 2, # dž
```

IJE.txt:

```
antitijelo  
avioprijevoznik  
besprijekoran  
besprijekorno  
besvijest
```

❖ Program za raščlambu hrvatskih riječi u slogove

❖ Demonstracija programa

Rezultat:

R: škripanje

S: škripanje, škropljenje, štrecanje, ždrebešće, ždrijebe, ždrijelo, smrekovina, stjuardesa, škrabica, škrapa, škrinja, posestrimstvo, dvije, higijena, znanstveni, sjevernjak, zlopatnik, predodžba, usmeni, sunašće, istarski, pašće, primjedba, jezerce, mladićstvo, bapski, postkoncilski, mlijeko, renesansni, sportski,

Antonio Jovanović,
student MathOSa



Algorithm 3 FILTERSYLLABLES(SYLLABLE_LIST, WEIGHTS, PREFIX_RULES, SUFFIX_RULES) filters syllables according to rules

```

1: Input syllable_list, weights, prefix_rules, suffix_rules
2: Output new_syllable_list
3: for syllable in syllable_list do
4:   prefix  $\leftarrow$  prefix of syllable
5:   suffix  $\leftarrow$  suffix of syllable
6:   if prefix in prefix_rules and suffix in suffix_rules then
7:     insert syllable into new_syllable_list
8: return new_syllable_list

```

Algorithm 4 SYLLABLES(WORD, WEIGHTS, PREFIX_RULES, SUFFIX_RULES) finds all syllable decompositions of a word

```

1: Input word, weights, prefix_rules, suffix_rules
2: Output syllable_list
3: word  $\leftarrow$  transform every 'ije' and 'je' to 'jat' accordingly
4: nucleus_list  $\leftarrow$  SYLLABLENUCLEI(WORD, WEIGHTS)
5: possible_syllables  $\leftarrow$  []
6: for i from 1 to length(word) do
7:   if nucleus_list[i] = True then
8:     possible  $\leftarrow$  POSSIBLESYLLABLES(word[i], word[:i], word[i + 1:], nucleus_list)
9:     append FILTERSYLLABLES(possible, weights, prefix_rules, suffix_rules) to possible_syllables
10: syllables  $\leftarrow$  list of every possible combination of syllables from possible_syllables that equals word
11: return syllables

```

❖ Adaptivni algoritmi s obzirom na ljestvice i slogovne granice

❖ Baza (≈ 5800 slogova)
prvoga hrvatskog slogovnika

5/5

