

## Hrvatski mrežni subčestični (strukturni) rječnik

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

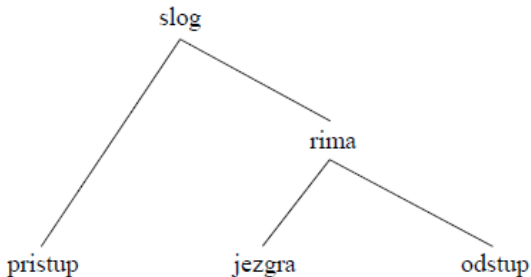
CLGF-2020

Osijek, 25.-26. rujna 2020.

# Sadržaj:

- 1 Nutarnja struktura riječi (Mario Essert)
  - Subčestice riječi
  - Slogotvorna pravila
  - On-line raščlamba
  
- 2 O algoritmu rastavljanja (Antonio Jovanović)
  - Jezgre
  - Pristupi i odstupi
  - Baza slogova u SSFu

# Model sloga



- Osnovni model sloga sliči apstraktnome stablu u koje se svi slogovi trebaju uklopiti kako bi bili prepoznati kao prihvatljivi slogovi pojedinoga jezika<sup>1</sup>
- Slog se može objasniti kao lûk sonornosti/zvonačnosti, odnosno kao fonološka jedinica sastavljena od jednog ili više odsječaka (glasova), koja ima jedan vrh sonornosti.<sup>2</sup>

<sup>1</sup>Jelaska, Zrinka (2004) *Fonološki opisi hrvatskoga jezika. Glasovi, slogovi, naglasci*. Zagreb: Hrvatska sveučilišna naklada.

<sup>2</sup>Škiljan, Dubravko (1980) *Pogled u lingvistiku*. Zagreb: Školska knjiga

# Model morfema<sup>3</sup>

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

<sup>3</sup>Marković, Ivan (2013) *Hrvatska morfonologija*. Zagreb: Disput

# Model silabomorfema<sup>4</sup>

## Infinitivni dočetcí (silabomorfemi) tipa *-v̄cv̄c(v)* i prezentski dočetcí (silabomorfemi) *-v̄cv̄c(v)*

(Užarević 2013: 229-231)

Dočetcí na <i>-a</i>		Dočetcí na <i>-e</i>		Dočetcí na <i>-i</i>		Dočetcí na <i>-o</i>		Dočetcí na <i>-r</i>		Dočetcí na <i>-u</i>	
Inf.	Prez.	Inf.	Prez.	Inf.	Prez.	Inf.	Prez.	Inf.	Prez.	Inf.	Prez.
-ābāt > -ābān		-ēbāt > -ēbān		-ībāt > -ībān		-ōbāt > -ōbān		-r̄bāt > -r̄bān		-ūbāt > -ūbān	
-ābīt > -ābīn		-----		-ībīt > -ībīn		-----		-r̄bīt > -r̄bīn		-ūbīt > -ūbīn	
-ābljāt > -ābljān		-----		-----		-----		-----		-----	
-ābrīt > -ābrīn		-----		-----		-----		-----		-----	
-ācāt > -ācān		-ēcāt > -ēcān		-īcāt > -īcēn		-ōcāt > -ōcān		-r̄cāt > -r̄cān		-ūcāt > -ūcān	
-ācīt > -ācīn		-----		-----		-ōcīt > -ōcīn		-----		-ūcīt > -ūcīn	
-āckāt > -āckān		-----		-----		-----		-----		-ūckāt > -ūckān	
-----		-----		-----		-----		-r̄cnīt > -r̄cnīn		-ūcnīt > -ūcnīn	
-āčāt > -āčān		-----		-īčāt > -īčēn		-----		-----		-ūčāt > -ūčān	
-āčīt > -āčīn		-ēčīt > -ēčīn		-īčīt > -īčīn		-----		-r̄čīt > -r̄čīn		-ūčīt > -ūčīn	
-ācāt > -ācān		-----		-īcāt > -īcān		-----		-----		-ūcīt > -ūcīn	
-----		-----		-īckāt > -īckān		-----		-r̄ckāt > -r̄ckān		-ūckāt > -ūckān	
-----		-----		-īcnīt > -īcnīn		-----		-----		-ūcnīt > -ūcnīn	
-ādāt > -ādān		-ēdāt > -ēdān > -ēdēm		-īdāt > -īdān		-ōdāt > -ōdān		-r̄dāt > -r̄dān		-----	
-ādīt > -ādīn		-ēdīt > -ēdīn		-īdīt > -īdīn		-ōdīt > -ōdīn		-r̄dīt > -r̄dīn		-ūdīt > -ūdīn	
-ādāt > -ādān		-ēdāt > -ēdān		-īdāt > -īdān		-----		-----		-----	

<sup>4</sup>Užarević, Josip, Skup u Osijeku: "Hrvatski jezik u računalnome jezikoslovlju",

# Fonologija - ljestvice zvanačnosti (zvonkosti, sonornosti)

## Sedam razina<sup>5</sup>

- ① šumnici ili pravi konsonanti: p, t, k, c, č, ć, s, š, f, h, b, d, g, đ, dž, z, ž
- ② zvonki usnici ili labijalni sonanti: m, v
- ③ svijetli nosnici ili nazali: n, nj
- ④ bočnici ili laterali: l, lj
- ⑤ titrajnik ili vibrant: r
- ⑥ prijelaznik ili poluvokal: j
- ⑦ samoglasnici ili vokali: a, e, i, o, u

## Četiri razine (*J. Užarević, 2019.*)

- ① p, t, k, c, č, ć, s, š, f, h
- ② b, d, g, đ, dž, z, ž
- ③ j, l, lj, m, n, nj, r, v
- ④ a, e, o, i, u, r, ije, je, i, u

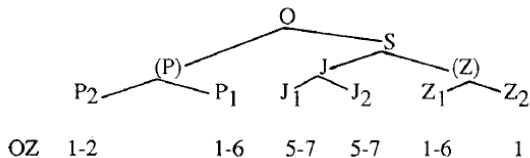
<sup>5</sup>Babić, Zrinka, Slogovna struktura hratskoga književnog jezika, *Jezik*, Vol. 36 No. 4, 1988

# Tri opća slogovna (slogotvorna) pravila

- ① Zakon otvorenih slogova: slogovi teže ostati otvorenima, tj. tendiraju da se završe samoglasnikom: si-na-go-ga, cr-kva (ne crk-va), go-ri-vo
- ② Zakon narastajuće (uzlazne) zvonkosti (zvučnosti): ako se slog sastoji od samoglasnika i više suglasnika, suglasnici se raspoređuju tako da idu od niže zvučnosti prema višoj: str-pljen (113/4-134.), zgro-mi-ti (2234-34-14), cr-kva (13/4-134), be-spo-mo-ćnost (24-114-13411). Pravilo uzlazne zvonkosti ne vrijedi za slogove na kraju riječi – ondje se može nalaziti jedan ili više suglasnika različite zvonkosti u različitim kombinacijama (a-dvent, cr-ven-kast, grozd) (Zakon otvorenih slogova zapravo je uključen u zakon narastajuće zvučnosti, jer su samoglasnici po odredbi najzvučniji među glasovima.)
- ③ Dopunsko pravilo (mehaničko, odnosno nefonološko): Zakon trećega/četvrtoga suglasnika: Prvi od tri uzastopna suglasnika (bilo zvučna, bilo bezvučna) može – prema potrebi – ići u prethodni slog: od-gri-sti (ali se kao početni ili kanonski oblik uzima o-dgri-sti), di-rek-tno (di-re-ktno). Ako imamo 4 uzastopna suglasnika, tada prva dva idu u prethodni slog, tj. razdioba je 2 + 2: gang-ster, a-dvent-ski

# Hrvatski slog sa 7-razinskom ljestvicom (Jelaska, 2004)

P-pristup, J-jezgra, Z-odstup, OZ-razina zvučnosti



- ① šumnici ili pravi konsontani: p, t, k, c, č, ć, s, š, f, h, b, d, g, đ, dž, z, ž
- ② zvonki usnici ili labijalni sonanti: m, v
- ③ svijetli nosnici ili nazali: n, nj
- ④ bočnici ili laterali: l, lj
- ⑤ titrajnik ili vibrant: r
- ⑥ prijelaznik ili poluvokal: j
- ⑦ samoglasnici ili vokali: a, e, i, o, u

## 13 razina (2019.)

- ① p, t, k, c, č, ć
- ② b, d, g, đ, dž
- ③ s, š, f, h
- ④ z, ž
- ⑤ v, m
- ⑥ n, nj
- ⑦ l, lj
- ⑧ r
- ⑨ j
- ⑩ i, u, r
- ⑪ ije, je
- ⑫ e, o
- ⑬ a



<http://m-brojevi.com/slogovi2/>

## Slogovi

rijeci.txt:

```
ispijen,svijen,trijem,grijem, smijem, sijem, dviju,
obzir,bezobziran,obzor,povlačiti,navlaka,uvlaka,
izbavljenje, predikativnost, aktivnost, slogovnik, stalnost,
daljnji, marni, stvarni, umni,šumni,
razumni,sublimni,svratiti,svraka,krajnji, praskozorje,
```

padine.py:

```
## Padine pristupa i odstupa ###
# \u24c7' = r (slogotvorno), \u01cc' = nj, \u01c9' = lj, \u01c6' = dž,
\u2c21' = ije, \u2c51' = je
pristupi = [
[
```

spis 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
```

Četiri tablična polja/datoteke (sve je moguće mijenjati i on-line):

- 1 **riječi.txt** - riječi (koje moraju biti odvojene razmakom ili zarezom).
- 2 **tezine.py** - ljestvice zvonačnosti, trenutno s 13 razina.
- 3 **padine.py** - dopuštene kombinacije glasova u pristupima i odstupima
- 4 **IJE.txt** - rječi s dugim 'jat-om' (leme) iz kojih se generiraju pojavnice

# Rezultati raščlambe

## Rezultat:

R: bezobziran

S: ['be.zo.bzi.ran', 'be.zo.bzir.an', 'be.zob.zi.ran', 'be.zob.zir.an', 'bez.o.bzi.ran', 'bez.o.bzir.an', 'bez.ob.zi.ran', 'bez.ob.zir.an']

R: obzor

S: ['o.bzor', 'ob.zor']

R: povlačiti

S: ['po.vla.či.ti', 'po.vla.čit.i', 'po.vla.či.ti', 'po.vla.čit.i', 'pov.la.či.ti', 'pov.la.čit.i', 'pov.la.či.ti', 'pov.la.čit.i']

R: navlaka

S: ['na.vla.ka', 'na.vlak.a', 'nav.la.ka', 'nav.lak.a']

R: uvlaka

S: ['u.vla.ka', 'u.vlak.a', 'uv.la.ka', 'uv.lak.a']

R: izbavljenje

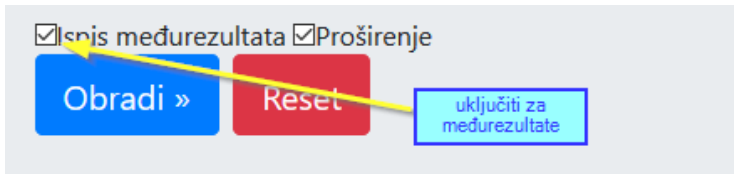
S: ['iz.ba.vlje.nje', 'iz.ba.vljenj.e', 'iz.bav.lje.nje', 'iz.bav.ljenj.e', 'iz.ba.vlje.nje', 'iz.ba.vljenj.e', 'iz.bav.lje.nje', 'iz.bav.ljenj.e']

R: predikativnost

S: ['pre.di.ka.ti.vnost', 'pre.di.ka.tiv.nost', 'pre.di.kat.i.vnost', 'pre.di.kat.iv.nost', 'pre.dik.a.ti.vnost', 'pre.dik.a.tiv.nost', 'pre.dik.at.i.vnost', 'pre.dik.at.iv.nost', 'pred.i.ka.ti.vnost', 'pred.i.ka.tiv.nost', 'pred.i.kat.i.vnost', 'pred.i.kat.iv.nost', 'pred.ika.a.ti.vnost', 'pred.ika.a.tiv.nost', 'pred.ikat.i.vnost', 'pred.ikat.iv.nost']

Sva dopuštenja rastavljanja po pravilima (padina i težina).  
Prva raščlambe u nizu predstavlja **kanonski oblik** raščlambe.

# Međurezultati - lako oblikovanje padina



R: predikativnost

pr\_\_e\_\_dikativnost

M: ('pre', 2, 0) 1.8.12. | ('pred', 2, 1) 1.8.12.2. |

F: ['pre', 'pred']

pred\_\_i\_\_kativnost

M: ('i', 0, 0) 10. | ('di', 1, 0) 2.10. | ('ik', 0, 1) 10.1. | ('dik', 1, 1) 2.10.1. |

F: ['i', 'di', 'ik', 'dik']

predik\_\_a\_\_tivnost

M: ('a', 0, 0) 13. | ('ka', 1, 0) 1.13. | ('at', 0, 1) 13.1. | ('kat', 1, 1) 1.13.1. |

F: ['a', 'ka', 'at', 'kat']

predikat\_\_i\_\_vnost

M: ('i', 0, 0) 10. | ('ti', 1, 0) 1.10. | ('iv', 0, 1) 10.5. | ('tiv', 1, 1) 1.10.5. | ('ivn', 0, 2) 10.5.6. | ('tivn', 1, 2) 1.10.5.6. |

F: ['i', 'ti', 'iv', 'tiv']

predikativn\_\_o\_\_st

M: ('ost', 0, 2) 12.3.1. | ('nost', 1, 2) 6.12.3.1. | ('vnost', 2, 2) 5.6.12.3.1. |

F: ['ost', 'nost', 'vnost']

S: ['pre.di.ka.ti.vnost', 'pre.di.ka.tiv.nost', 'pre.di.kat.i.vnost', 'pre.di.kat.iv.nost', 'pre.dika.ti.vnost', 'pre.dika.tiv.nost', 'pre.dik.at.i.vnost', 'pre.dik.at.iv.nost', 'pred.i.ka.ti.vnost', 'pred.i.ka.tiv.nost', 'pred.i.kat.i.vnost', 'pred.i.kat.iv.nost', 'pred.ika.ti.vnost', 'pred.ika.tiv.nost', 'pred.ik.at.i.vnost', 'pred.ik.at.iv.nost']

# ALGORITAM RASTAVLJANJA I IMPLEMENTACIJA U SSF-u

# Traženje jezgri slogova u riječi

---

**Algorithm 1** SYLLABLENUCLEI(*WORD*, *WEIGHTS*) finds all syllable nuclei in a word

---

```

1: Input word, weights
2: Output list of boolean values, i-th value is True if word(i) is a nucleus, False otherwise
3: nucleus_list  $\leftarrow$  []
4: n  $\leftarrow$  length(word)
5: vowels  $\leftarrow$  ['a', 'e', 'i', 'o', 'u']
6: for i = 1 to n do
7:   sound  $\leftarrow$  word[i]
8:   cond_1  $\leftarrow$  sound is in vowel or sound is jat
9:   cond_2  $\leftarrow$  sound = 'r' and i = 1 and weights of following sound is  $\leq$  7
10:  cond_3  $\leftarrow$  sound = 'r' and i = n and weights of preceding sound is  $\leq$  7
11:  cond_4  $\leftarrow$  sound = 'r' and i not in [1,n] and weights of neighbouring sounds is  $\leq$  7
12:  cond_5  $\leftarrow$  sound = 'l' and i = n and weights of preceding sound is  $\leq$  6
13:  cond_6  $\leftarrow$  sound in ['l', 'lj'] and i not in [1,n] and weights of neighbouring sounds is  $\leq$  6
14:  cond_7  $\leftarrow$  sound = 'm' and i not in [1,n] and weights of neighbouring sounds is  $\leq$  4
15:  cond_8  $\leftarrow$  sound in ['n', 'nj'] and i not in [1,n] and weights of neighbouring sounds is  $\leq$  5
16:  if cond_1 or cond_2 or cond_3 or cond_4 or cond_5 or cond_6 or cond_7 or cond_8 then
17:    append True to nucleus_list
18:  else
19:    append False to nucleus_list
20: return nucleus_list

```

---

## Rezultat:

R: drvo

S: ['dr.vo', 'drv.o']

R: trljati

S: ['tr.lja.ti', 'tr.ljat.i', 'trlj.a.ti', 'trlj.at.i']

R: rđati

S: ['r.đa.ti', 'r.đat.i', 'rđ.a.ti', 'rđ.at.i']

R: rvač

S: ['r.vač', 'rv.ač']

R: rzati

S: ['r.za.ti', 'r.zat.i', 'rz.a.ti', 'rz.at.i']

R: bicikl

S: ['bi.ci.kl', 'bi.cik.l', 'bic.i.kl', 'bic.ik.l']

R: monokl

S: ['mo.no.kl', 'mo.nok.l', 'mon.o.kl', 'mon.ok.l']

R: vltava

S: ['vl.ta.va', 'vl.tav.a', 'vlt.a.va', 'vlt.av.a']

# Svi mogući ispravni slogovi u riječi

**Algorithm 2** POSSIBLESYLLABLES(NUCLEUS, PREFIX, SUFFIX, NUCLEUS\_LIST) finds all possible syllables in a word that contain nucleus

---

```

1: Input nucleus, prefix, suffix, nucleus_list
2: Output List of all possible syllables containing nucleus
3: syllable_list  $\leftarrow$  [nucleus]
4: if there are no nuclei in prefix then
5:   append prefix + nucleus to syllable_list
6: if there are no nuclei in suffix then
7:   if there are no nuclei in prefix then
8:     append prefix + nucleus + suffix to syllable_list
9:   else
10:    append prefix + nucleus to syllable_list
11: if there are nuclei in prefix then
12:   bottom  $\leftarrow$  -1
13:   for i  $\leftarrow$  length(prefix) to 1 do
14:     if nucleus_list[i] = True then
15:       break
16:     else
17:       if bottom = -1 and weight[prefix[i]] < weight[prefix[i - 1]] then
18:         bottom  $\leftarrow$  i
19:       else if bottom  $\neq$  -1 then
20:         if bottom < length(prefix) then
21:           break
22:         if bottom = length(prefix) and weight[prefix[i]] > weight[prefix[i + 1]] then
23:           break
24:       append prefix[i :] + nucleus to syllable_list

```

```
25: if there are nuclei in suffix then
26:   bottom  $\leftarrow$  -1
27:   for i  $\leftarrow$  1 to length(suffix) do
28:     if nucleus_list[i + length(prefix) + 1] = True then
29:       break
30:     else
31:       if bottom = -1 and weight[suffix[i]] < weight[suffix[i + 1]] then
32:         bottom  $\leftarrow$  i
33:       else if bottom  $\neq$  -1 then
34:         if bottom > 1 then
35:           break
36:         if bottom = 1 and weight[suffix[i]] < weight[suffix[i - 1]] then
37:           break
38:       append syllable + suffix[i] to syllable_list for every syllable in syllable_list that ends in suffix[i - 1]
       or nucleus if i = 0
39: return syllable_list
```

---



---

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

```

---

# Postavke za pretragu u SSF-u

Riječi

T-Strukture

Rečenice

MSY

LEX

MWE



## Postavke pretrage

Broj morfova u bazi: 4.418 | Broj slogova u bazi: 7.591 | Broj silabomorfema u bazi: 323.253

Maksimalni broj rezultata:  \*0 za neograničeno

### Tip pretrage:

Morfovi  Slogovi  Silabomorfemi

### WOS/SOW filter:

[Označi sve](#) [Odznači sve](#)

- Vrsta riječi
  - Imenica*
  - Zamjenica
  - Pridjev

[Označi sve](#) [Odznači sve](#)

- GPS
- Lokalitet
- Definicija
- Primjer

### Pozicija u riječi

Bilo koja  01  02  03  04  05  06  07  08  09  10  11  Zadnji

MSY

- odrediti obilježja riječi koje se raščlanjuju
- odrediti mjesto sloga po abecedi unutar riječi

# Rezultati u SSF-u

[← Povratak](#)

A B C Č Ć D DŽ Đ E F G H I J **K** L LJ M N NJ O P R S Š T U V Z Ž

Postoji 496 riječi koje zadovoljavaju kriterij pretrage. Prikazuje se prvih 100

## KA

Riječi:

pen-ka-la lo-ka-ci-ja stru-ka bilj-ka šija-ka broj-ka tra-ka sli-ka bru-ka blo-ka-da lu-ka-vost bolj-ka do-ka-zi-va-nje vo-ka-ci-ja  
 bu-ka zvu-ka str-ka alj-ka-vost spoj-ka mu-ka ro-ka pri-ka-za sto-ka pri-ka-zi-va-nje var-ka manj-ka-vost luc-ka-stost školj-ka  
 ko-ka ru-ka-vac de-ka de-ka-gram fle-ka stan-ka zbir-ka snim-ka kli-ka vo-ka-li-za-ci-ja po-ka-za-telj po-ka-zi-vač po-ka-zi-va-nje  
 u-ka-zi-va-nje žen-ka se-ka stran-ka-ma la-ka ru-ka ša-ka la-ka-ta ru-ka-ma ba-ka zra-ka maj-ka ri-ka do-ka-ze bar-ka  
 rije-ka koc-ka-ma do-ka-zi-va-nja član-ka stran-ka šmin-ka lije-ka pla-ka-te pla-ka-ta li-ka ...

## KAC

Riječi:

ku-kac šo-kac džu-kac **tu-kac**

tukac

## KAČ

Riječi:

ska-kač š  
 WOS: Slogovi Vrsta riječi • Imenica  
 Rod • Muški Broj • Jedinina  
 Padež • Nominativ

ač pe-kač so-kač ba-kač mi-kač

# Hvala na pozornosti



Nema više (kraj je blizu) 😊!