

UDK 811.163.6:004.42:681.3

Melita Zemljak, Pedagoška fakulteta v Mariboru

Zdravko Kačič, Fakulteta za elektrotehniko, računalništvo in informatiko v Mariboru

Simon Dobrišek, Fakulteta za elektrotehniko v Ljubljani

Jerneja Gros, Fakulteta za elektrotehniko v Ljubljani

Peter Weiss, Inštitut za slovenski jezik Frana Ramovša v Ljubljani

RAČUNALNIŠKI SIMBOLNI FONETIČNI ZAPIS SLOVENSKEGA GOVORA

V slovenskem jezikovnem prostoru se je sredi 90. let pojavila težnja po enotnem računalniškem zapisu slovenskega govora. Ta bi standardiziral in poenotil raziskovalno delo na področju govornih in jezikovnih tehnologij, ki se nanaša na fonetične označitve slovenskih govornih besedil. Sodelovanje mariborske in ljubljanske skupine raziskovalcev (Fakulteta za elektrotehniko, računalništvo in informatiko, Pedagoška fakulteta; Fakulteta za elektrotehniko, ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša) je rodilo računalniški fonetični zapis simbolov IPA in MRPA. Predvsem slednji so velikega pomena za enostavno in prijazno računalniško uporabo.

In the Slovene-speaking area in the mid-1990's there appeared a trend towards a unified computer transcription of Slovene speech. This transcription would standardize and unify research in the field of speech and language technologies that touches upon the phonetic symbolization of Slovene spoken texts. The cooperation of research groups from Maribor and Ljubljana (Faculty of Electrical Engineering and Computer Science; Faculty of Education; Faculty of Electrical Engineering; The Scientific Research Centre of SASA, Fran Ramovš Institute of the Slovenian Language) resulted in the development of computer-based phonetic transcriptions of the symbols in IPA and MRPA. Largely the latter are of great importance for simple and user-friendly computer usage.

Ključne besede: slovenski jezik, fonetika, računalniška fonetična transkripcija, simbol MRPA (machine readable alphabet), simboli IPA (international phonetic alphabet)

Key words: Slovenian language, phonetics, computer-based phonetic transcription, MRPA (machine readable alphabet) symbols, IPA (International Phonetic Alphabet) symbols

1 Uvod

Namen avtorjev računalniškega fonetičnega zapisa slovenskega govora je izbor in predstavitev računalniških simbolov v obliki posebne abecede. Tako bi standardizirali in poenotili raziskovalno delo na področju govornih in jezikovnih tehnologij, ki se nanaša na fonetične označitve slovenskih govornih besedil. Prva pobuda za standardizacijo je bila predstavljena na Elektrotehniški in računalniški konferenci v Portorožu. V pisni obliki je bila objavljena v zborniku te konference – ERK '95 (Dobrišek in dr. 1995: 247–250). Ker na področju jezikovnih in govornih tehnologij v svetu potekajo obširne raziskave, ki so v boju za ohranitev jezikovne identitete in jezikovnega napredka posameznega jezika navdse pomembne tudi za Slovence, je za tako maloštevilen narod vsekakor bistveno, da pride do standardizacije računalniškega simbolnega fonetičnega zapisa slovenskega govora, ki bi tako ohranjal narodno in raziskovalno enotnost in nas s tem postavil ob bok svetovni eliti. V preteklih letih je bilo v slovenskem prostoru opravljenih več projektov izdelave infrastrukture – zbir izgovorjav (SNABI, SPEECHDAT, Onomastica na Fakulteti za elektrotehniko,

računalništvo in informatiko v Mariboru, zbirka GOPOLIS na Fakulteti za elektrotehniko v Ljubljani). Od tod torej nuja po standardizaciji za poenotenje nadaljnjega dela.

Opredelitev do glasoslovnih pojmov, označevanje razčlenjenega govora na govorne enote na več ravni zaradi večje splošnosti so predstavljeni že v Pobudi za standardizacijo /.../. (Gl. prejšnjo op.) Dne 13. 2. 1997 so bili predlogi skupno predstavljeni in ovrednoteni s strani Fakultete za elektrotehniko, računalništvo in informatiko Univerze v Mariboru (Laboratorij za digitalno procesiranje signalov) ter Fakultete za elektrotehniko Univerze v Ljubljani (Laboratorij za umetno zaznavanje).¹

2 Računalniški fonetični simboli

Za označevanje govora na ožji fonetični ravni je standardna uporaba mednarodne fonetične abecede IPA (Fourcin in dr. 1989; Šuštaršič, Komar, Petek 1995: 86–90), za lažjo preslikavo teh posebnih fonetičnih simbolov v standardne (ASCII/ANSI) fonetične simbole pa uporabljamo mednarodno uveljavljene simbole abecede MRPA (machine readable alphabet), ki jim abeceda SAMPA najpogosteje služi kot osnova.²

V naslednjih podpoglavjih so predstavljeni posamezni fonetični simboli, zapisani najprej po abecedi MRPA, ki ji sledijo simboli IPA. Tema zapisoma so dodani primeri besed, v katerih se ti simboli pojavljajo, in sicer v ortografski transkripciji in fonetični transkripciji MRPA z dodano angleško pomensko ustreznico. V fonetični transkripciji MRPA so dolgi glasovi označeni z dvopičjem (:). Simbol " označuje začetek naglašene zloga (IPA ').

Delitev simbolov na posamezne skupine se naslanja na podobne delitve glasov³ slovenskega knjižnega jezika (*Slovenski pravopis* 2001; Toporišič 1958/59, 1962, 1967, 2000). Razdeljeni so po odprtostni stopnji govornega trakta – od največje do najmanjše. Pri soglasnikih so zabeležene tudi skupine simbolov za alofone, katerih uporaba ali neuporaba za znanstveno raziskovanje je odvisna od želene natančnosti rezultatov.

2.1 Samoglasniki

Slovenski knjižni jezik pozna osem samoglasnikov (glasov največje odprtostne stopnje): i, e, E, a, O, o, u in ∂. Kvantitativno so razdeljeni na dolge (teh je 7) in kratke samoglasnike (12); od teh so prvi samo naglašeni, drugi pa naglašeni (6) ali nenaglašeni (6).

¹ Jezikoslovne nasvete sta prispevala še predstavnik Filozofske fakultete v Ljubljani (dr. Primož Vitez, doc. dr. Smiljana Komar), v dobršni meri pa so bili upoštevani tudi nasveti prof. dr. Jožeta Toporišiča. Vsem se za pomoč zahvaljujemo.

² Abeceda SAMPA je bila definirana v okviru projekta SAM – SAM phonetic alphabet.

³ Izraz *glas* pomeni akustično realizacijo fonema.

2.1.1 Dolgi samoglasniki

Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
i:	i:	pila	"pi:la	file
e:	e:	pet	"pe:t	five
E:	ɛ:	teta	"tɛ:ta	aunt
a:	a:	mama	"ma:ma	mother
O:	ɔ:	voda	"vO:da	water
o:	o:	pot	"po:t	path
u:	u:	suša	"su:Sa	drought

2.1.2 Kratki samoglasniki

– naglašeni:

Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
i	i	sit	"sit	full, satiated
E	ɛ	zelen	zɛ"lɛn	green
a	a	čas	"tSas	time
O	ɔ	potop	pO"tOp	flood
u	u	kruh	"krux	bread
@	ə	tema, vrt	"t@ma, "v@rt	darkness, garden

– nenaglašeni:

Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
i	i	prijava	pri"ja:va	application
E	ɛ	zelen	zɛ"lɛn	green
a	a	tema	"t@ma	darkness
O	ɔ	potop	pO"tOp	flood
u	u	konzul	"ko:nzul	consul
@	ə	veter, vrtiček	"ve:t@r, v@r"ti:č@k	wind, small garden

2.2 Soglasniki

Soglasniški sistem knjižne slovenščine vsebuje šest zvočnikov, šest zapornikov, šest pripornikov in tri zlitnike.

2.2.1 Zvočniki

V slovenskem knjižnem jeziku zvočniki (glasovi srednje odprtostne stopnje) ne vplivajo na zvonečnost nezvočnikov pred seboj ali za seboj. Zvočniki tudi nimajo enakovrednih oz. ustreznih nezvonečih parov (edini izjemi sta [w]⁴ in [W]).⁵

Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
m	m	miza	"mi:za	table
n	n	noč	"no:tS	night
l	l	leto	"le:tO	year
r	r	riba	"ri:ba	fish
v	v	voda	"vO:da	water
j	j	jeza	"je:za	anger

Nekateri izmed zvočniških fonemov ločijo alofone:

- alofon fonema /m/ je: [F] pred /f/ in/ali /v/;
- alofoni fonema /n/ so: [N] pred /k/, /g/, /x/,
[F] pred /f/, /v/,
[n'] pred soglasnikom in na koncu besede;
- alofona fonema /l/ sta: [l'] pred soglasnikom in na koncu besede,
[U] za samoglasnikom in hkrati na koncu besede,
ali pred soglasnikom;
- alofoni fonema /v/ so: [w] pred zvonečimi soglasniki ne za samoglasnikom,
[W] pred nezvonečimi soglasniki ne za samoglasnikom,
[U] za samoglasnikom in hkrati na koncu besede ali pred soglasnikom;
- alofon fonema /j/ je: [I] za samoglasnikom in hkrati na koncu besede ali pred soglasnikom.

⁴ Fonemi so v spremnem besedilu ob preglednicah zapisani v poševnih oklepajih, npr. /v/, alofoni pa v oglatih, npr. [w].

⁵ Fonetični simboli, zapisani v spremnem besedilu ob preglednicah, so zaradi enostavnosti zapisa v celotnem prispevku zabeleženi s simboli MRPA.

Simbol fonema po MRPA	Simbol fonema po IPA	Simbol alofona po MRPA	Simbol alofona po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
m	m	F	ɱ	nimfa, sem vesela	"ni:Ffa, s@F vE"se:la	nymph, I am happy
n	n	N	ŋ	Anka, angel, Anhovo	"a:Nka, "a:NgEl, "a:Nxovo	Anka, angel, Anhovo
		F	ɱ	konfin, konvoj	koF"fi:n, koF"vo:l	boundary stone, convoy
		n'	n ^j	konj, konjski	"kOn', "ko:n'ski/ "kO:n'ski	horse, horsy
l	l	l'	lj	polj, poljski	"po:l', "po:l'ski	field, agrarian
l	l	U	u za soglasni- kom	cev, cevka tkal, tkalca	"ce:U, "ce:Uka "tka:U, "tka:Utsa	tube, small tube, he wove, weaver's
v	v	w	w	vzeti	"wze:ti	to take
v	v	W	ɱ	vsak	"Wsa:k	every
j	j	I	ɪ	lajna	"la:Ina	barrel-organ

2.2.2 Nezvočniki

Slovenski knjižni jezik deli nezvočnike (glasove najmanjše odprtostne stopnje) glede na način tvorbe na naslednje skupine glasov:

2.2.2.1 Nezvočniki glede na način tvorbe

2.2.2.1.1 Zaporniki

Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
p	p	pipa	"pi:pa	faucet
b	b	beda	"be:da	misery
t	t	teta	"tE:ta	aunt
d	d	delo	"de:lO	work
k	k	kolo	kO"lo:	wheel
g	g	goba	"go:ba	mushroom

Nekateri izmed zapornikov (fonemov) imajo v določenih položajih svoje alofone:

- alofona fonema /p/ sta: [p_n] pred /m/,
[p_f] pred /f/, /v/;
- alofona fonema /b/ sta: [b_n] pred /m/,
[b_f] pred /f/, /v/;

- alofona fonema /t/ sta: [t_l] pred /l/,
[t_n] pred /n/;
- alofona fonema /d/ sta: [d_l] pred /l/,
[d_n] pred /n/.

Simbol fonema po MRPA	Simbol fonema po IPA	Simbol alofona po MRPA	Simbol alofona po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
p	p	p_n	p̃	zob me boli	"zo:p_n mE bO"li:	I have a toothache
p	p	p_f	p ^f	ob fantu, rob vaze	Op_f "fa:ntu, "ro:p_f "va:zE	near a boy, the edge of a vase
b	b	b_n	b̃	z zobmi	z zOb_n"mi:	with teeth
b	b	b_f	b ^f	obveza	Ob_f"ve:za	bandage
t	t	t_l	t ^l	tla	"t_lla	floor
		t_n	t̃	tnalo	"t_nna:lO	chopping block
d	d	d_l	d ^l	dleto	"d_lle:tO	chisel
		d_n	d̃	dno	"d_nnO	bottom

2.2.2.1.2 Priporniki

Slovenski knjižni priporniki so: /f/, /s/, /z/, /S/, /Z/ in /x/. Fonem /x/ ima alofon [G].

Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
f	f	figa	"fi:ga	fig
s	s	soba	"sO:ba	room
z	z	zima	"zi:ma	winter
S	ʃ	šoba	"So:ba	pout
Z	ʒ	žoga	"Zo:ga	ball
x	x	hiša	"xi:Sa	house
G	ɣ	h gori	G "gO:ri	to the mountain

2.2.2.1.3 Zlitniki

Slovenski knjižni zlitniki so /ts/, /tS/ in /dZ/.

Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
ts	ts	cula	"tsu:la	bundle
tS	tʃ	čelo	"tSE:lO	forehead
dZ1	dʒ	džungla	"dZu:Ngla	jungle
dz2	dz	Kocbek	"ko:dzbEk	(surname) Kocbek

2.2.2.2 Nezvočniki glede na zvonečnost

Vse tri skupine nezvočnikov vsebujejo tako zvoneče kot nezvoneče glasove najmanjše odprtostne stopnje. Zvonečnost nezvočnikov se v slovenskem knjižnem jeziku spreminja glede na zvonečnost/nezvonečnost sosednjih fonemov, ki so bodisi v besedni bodisi v govorni enoti.⁸

2.2.2.2.1 Nezvoneči nezvočniki

Spremembe nezvonečih nezvočnikov /p, t, k, f, s, S, x/ so enake v besedni in v govorni enoti. Nezvoneči fonemi ostanejo nezvoneči v vseh položajih, razen pred zvonečimi nezvočniki. V tem položaju postanejo nezvoneči nezvočniki /p, t, s, S, k/ (nevtralizirani) zvoneči fonemi /b, d, z, Z, g/. Fonemi /ts, tS, f, x/ imajo alofone [dz, dZ, v, G].

Besedna enota				
Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
t > d	t > d	svatba	"sva:d̥ba	wedding
s > z	s > z	glasba	"gla:zba	music
S > Z	ʃ > ʒ	izvršba	iz"v@rZba	execution
k > g	k > g	prerokba	prE"ro:gba	prophecy
ts > dz	ts > dz	Kocbek	"ko:d̥zbEk	(surname) Kocbek
tS > dZ	tʃ > dʒ	enačba	E"na:d̥Zba	equation
f > v	f > v	Afganistan	av"ga:nistan	Afghanistan
x > G	x > ɣ	h gori	G "gO:ri	to the mountain

⁶ [dZ] je lahko tudi alofon fonema /tS/ pred zvonečim nezvočnikom: odločba – [O " dd_lo:dZba].

⁷ [dz] je alofon fonema /ts/ pred zvonečim nezvočnikom.

⁸ Izraz besedna enota se nanaša na sestavljene besede ter na predložne in predpanske zveze, govorna enota pa na vse druge zveze. Gl. Toporišič 1957/58: 27.

Simbol fonema po MRPA	Simbol fonema po IPA	Govorna enota		
		Ortografski zapis	Fonetični prepis	Angleška ustreznica
p > b	p > b	snop daj	"snOb "daj	give the sheaf
t > d	t > d	tat dobi	"ta:d "gre:/"grE	thief goes
s > z	s > z	glas doni	"gla:z dO"ni:	voice sounds
S > Z	ʃ > ʒ	iščēš žogo	"i:StSEZ "Zo:gO	you search for ball
k > g	k > g	vsak dan	"Wsa:g "da:n	every day
ts > dz	ts > dz	stric ga je	"stri:dz ga jE	uncle did this to him
tS > dZ	tʃ > dʒ	proč gre	"prOdZ"grE	he/she goes away
f > v	f > v	škof ga je	"SkOv ga jE	bishop did this to him
X > G	x > ɣ	strah ga je	s"tra:G ga jE	he is afraid

2.2.2.2.2 Zveneči nezvočniki

Spremembe zvenečih nezvočnikov se v primerjavi z nezvenečimi nezvočniki deloma razlikujejo glede na položaj, torej ali nastopajo v besedni ali govorni enoti.

Zveneči nezvočniki ostanejo v besedni enoti zveneči pred (1a) samoglasniki, (1b) zvočniki in (1c) zvenečimi nezvočniki. V govorni enoti prihaja do izgube zvenečnosti pred samoglasniki, zvočniki, nezvenečimi nezvočniki in v izglasju – zveneči nezvočniki ostanejo zveneči le pred (2) zvenečimi nezvočniki.

1a

Simbol fonema po MRPA	Simbol fonema po IPA	Besedna enota pred samoglasniki		
		Ortografski zapis	Fonetični prepis	Angleška ustreznica
b	b	biba	"bi:ba	bug
d	d	dedi	"de:di	grandfather
g	g	gol	"go:l	goal
z	z	zet	"zEt	son-in-law
Z	ʒ	žaga	"Za:ga	saw
dZ	dʒ	džezva	"dZe:zva	coffeepot

(1b)

		Besedna enota pred zvočniki		
Simbol fonema po MRPA	Simbol fonema po IPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica
b	b	objem	O"bjEm	embrace
d	d	dedje	"de:djE	grandfathers
g	g	glava	"gla:va	head
z	z	zrelost	"zre:lOst	maturity
Z	ʒ	lažna	"la:Zna	she is untrue

(1c)

		Besedna enota pred nezvočniki				
Simbol fonema po MRPA	Nezveneči			Zveneči		
	Ortografski zapis	Fonetični prepis	Angleška ustreznica	Ortografski zapis	Fonetični prepis	Angleška ustreznica
b	zobca	"zo:ptsɑ	dent, tooth	obdati, ob dedu	Ob"da:ti, Ob "de:du	to surround, by grandfather
d	godčev	"go:ttSEU	fiddler's	odgnati, od gore	Od"ɡna:ti, Od "ɡO:rE	to chase away, from mountain
g	bogcu	"bo:ktsu	to god	Bogdan	"bo:ɡdan	(first name) Bogdan
z	grozd	"grOst	bunch of grapes	zdaj, čez deda	"zdaj, tSEz "de:da	now, over grandfather
Z	Nežka	"ne:Ska	(first name) Nežka	žgati	"Zga:ti	to burn

		Govorna enota pred zvonečimi nezvočniki		
Simbol fonema po MRPA	Ortografski zapis	Fonetični prepis	Angleška ustreznica	
b	golob gre	ɡO"lo:b "gre:/ ɡO"lo:b "grE	pigeon goes	
d	divjad gre	diw"ja:d "gre:/ diw"ja:d "grE	game goes	
g	rog zveni	"ro:ɡ zve"ni:	horn sounds	
z	rez boli	"re:z bO"li:	cut hurts	
Z	gož gre	"go:Z "gre:/ "go:Z "grE	aesculapian snake goes	
dZ	govoreč glas	ɡOvO"rEdZ "gla:s/ ɡOvO"re:dZ "gla:s	speaking voice	

3 Sklep

V prispevku je podan računalniški simbolni zapis slovenskega govora, ki je plod sodelovanja mariborske (Fakulteta za elektrotehniko, računalništvo in informatiko, Pedagoška fakulteta) in ljubljanske (Fakulteta za elektrotehniko, ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša) skupine raziskovalcev s področja avtomatskega razpoznavanja in sinteze slovenskega govora. Podani so simboli tako zapisov MRPA kot zapisov IPA, od katerih so predvsem prvi velikega pomena za preprosto in prijazno računalniško uporabo. Pri soglasnikih je podan podrobnejši prikaz simbolov, kot ga obe raziskovalni skupini uporabljata danes, pokrije pa pravzaprav celotni sistem slovenskega (knjižnega) jezika. Želja vseh, ki smo kakor koli sodelovali pri oblikovanju računalniškega simbolnega fonetičnega zapisa slovenskega govora, pa je omogočanje enotnosti pri nadaljnjih raziskovanjih tega področja v že tako majhnem slovenskem prostoru, kar pomeni učinkovitejšo uveljavitev v svetu.

LITERATURA

- S. DOBRIŠEK, J. GROS, B. HORVAT, Z. KAČIČ, F. MIHELICH, 1995: Pobuda za standardizacijo računalniškega simbolnega fonetičnega zapisa slovenskega govora. *Zbornik ERK '95*. Portorož. 247–250.
- A. FOURCIN, G. HARLAND, W. BARRY, V. HAZAN, 1989: *Speech Input and Output Assessment, Multilingual Methods and Standards*. New York–Chichester–Brisbane–Toronto: Ellis Horwood Limited, John Wiley & Sons.
- B. PETEK, R. ŠUŠTARŠIČ, S. KOMAR, 1996: An Acoustic Analysis of Contemporary Vowels of the Standard Slovenian Language. *Proceedings ICSLP 96*. Wilmington (DE). 133–136.
- R. ŠUŠTARŠIČ, S. KOMAR, B. PETEK, 1995: Slovene: illustrations of the IPA. *Journal of the International Phonetic Association* 25/2. 86–90.
- R. ŠUŠTARŠIČ, S. KOMAR, B. PETEK, 1999: Slovene. *Handbook of the International Phonetic Association: A Guide to the Use of International Phonetic Alphabet*, Cambridge–New York–Melbourne–Madrid: Cambridge University Press. 135–139.
- J. TOPORIŠIČ, 1957/58: Sistemske premene soglasnikov v knjižnem govoru. *Jezik in slovstvo*. 21–29.
- – 1958/59: Alternativni soglasniški sklopi slovenskega knjižnega jezika. *Jezik in slovstvo*. 203–207.
- – 1962: Fonetika, fonologija in pravorečje. *Jezik in slovstvo*. 138–143.
- – 1967: Strukturiranost slovenskih glasov in predvidljivost njihove razvrstitve. *Jezik in slovstvo*. 92–96.
- – 1992: *Enciklopedija slovenskega jezika*. Ljubljana: Cankarjeva založba.
- – 2000: *Slovenska slovnica*. Maribor: Založba Obzorja. 37–100.
- Slovenski pravopis*. 2001. Ljubljana: SAZU, ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša. 69–86, 135–143.

SUMMARY

The paper gives the computer-based symbols for Slovene speech that was developed in cooperation between groups from Maribor (Faculty of Electrical Engineering and Computer Science, Faculty of Education) and Ljubljana (Faculty of Electrical Engineering, The Scientific Research Centre of SASA, Fran Ramovš Institute of the Slovenian Language) involved in research in the field of automatic speech recognition and Slovene speech synthesis. The transcription systems of MRPA as well as IPA are given, the former being of great importance for simple and user-friendly computer usage. For consonants a more detailed presentation of the symbols is given, as used today by both research groups; this in fact covers the entire system of (standard) Slovene. All of those who have worked on the development of a computerized symbolic phonetic transcription of Slovene speech strive for a unified system in further research in the field, particularly in light of the small Slovene-speaking area, as this means a more efficient promotion in the world.