
UDK 808.62:003

003.35

Stručni članak

Prihvaćeno 14.05.1998.

Juraj Bakran

Filozofski fakultet, Zagreb, Hrvatska

Damir Horga

Filozofski fakultet, Zagreb, Hrvatska

SAMPA ZA HRVATSKI

SAŽETAK

U članku se opisuje prijedlog skupa kompjutorski čitljivih ASCII znakova koji bi trebali zamijeniti standardne IPA simbole za fonetsku transkripciju hrvatskoga govora. Taj je prijedlog sastavljen na temelju načela SAMPA (Speech Assessment Methods Phonetic Alphabet) koji su objavili Wells et al. (1992). Osnovna, hrvatska verzija dostupna je na Internetu i podložna je izmjenama i dopunama.

Ključne riječi: fonetska transkripcija, SAMPA, hrvatski jezik

Standardna fonetska transkripcija koju predstavlja IPA (International Phonetic Alphabet) godinama je dopunjavana i usavršavana (vidi IPA, 1993). Mnogi se lingvisti i fonetičari koriste tom transkripcijom. Zapisivanje IPA znakova uvijek je bilo problem jer na standardnim pisačim strojevima ne postoji velik broj IPA znakova. Korisnici su se snalazili na razne načine, ručno nadopunjujući postojeći repertoar znakova. U posljednje vrijeme, kompjutorskog editiranja oblikovani su posebni *fontovi*, koji oslikavaju IPA znakovlje.

Međutim, problem se nanovo postavlja ako se želi koristiti samo skup ASCII znakova. Za to postoje barem dva jaka razloga: automatska obrada podataka i prijenos podataka elektroničkom poštom. Ako se ASCII datoteka (DOS tekst datoteka) šalje elektroničkom poštom, sigurno će se prenijeti samo između znakovi pod brojem 32 do 126. To su znakovi koji uključuju velika slova A ...Z, mala slova a...z, znamenke 0...9, znakovi: ! " ' () , - . / : ; { } [] \$ % & * + > < = _ | ` ~ ^ i razmak. Prema tome, želimo li se koristiti elektronskom poštom, fonetske znakove treba prekodirati služeći se samo tim skupom znakova.

Osnovni prijedlog za takvo prekodiranje nastao je između 1988. i 1991. godine. Autori su stručnjaci za govor iz 9 europskih zemalja koji su radili u sklopu SAM projekta (Speech Assessment Methods) (Wells et al. 1992). Sustav prekodiranja IPA nazvan je zbog toga SAMPA (od SAM Phonetic Alphabet). Svrha SAMPA prijedloga jest dakle oblikovati standardni fonetski alfabet koji bi bio strojno čitljiv i koji bi time poboljšao internacionalnu suradnju u istraživanjima govora.

Tako postoji zamjena za svaki od IPA simbola. Simboli se odabiru prema načelu sličnosti s IPA simbolima.

Na Internetu je dostupan popis SAMPA simbola za većinu europskih jezika. Otvori li se osnovna stranica:

<http://www.phon.ucl.ac.uk/home/sampa/home>

na njoj se mogu naći veze na druge tekstove i popise SAMPA simbola za razne jezike. Prema načelima SAMPA za hrvatski su Bakran i Horga (1997) predložili skup simbola koji obuhvaćaju fonemski repertoar i glavne alofone.

Oni IPA simboli koji su se poklapali s ASCII simbolima ostali su takvi i u SAMPA za hrvatski. Frikativ /š/ bilježi se za razliku od /s/ velikim slovom "S". Slično tomu, veliko slovo "Z" označava /ž/. Afrikate se pišu kao kombinacija "t" (za bezvučne) ili "d" (za zvučne) s jednim od frikativa, pa se sastoje od dva ili tri ASCII znaka. Tako /c/ bilježimo "ts", /č/ postaje "tS" a palatalizirana varijanta označava se jednostrukim apostroфом ' , pa /ć/ postaje "tS' ". Kao kod IPA, za /h/ zadržan je "x". Umjesto lambda, za /lj/ je predloženo "L" a za /nj/ predloženo je "J". Za sonant /v/ zadržan je simbol ali mu je dodana lijeva kosa zagrada (backslash), što znači da to nije kao u drugim jezicima zvučni frikativ.

Predložene su osim osnovnih simbola i transkripcije najvažnijih alofona. Ozvučeno /f/ bilježi se "v". Ozvučeno /c/, kao u /zec bi/, bilježi se "dz". Ozvučeno /h/ kao u /dah bi/ označava se s "G". Palataliziranost (frikativima)

označava se dodavanjem apostrofa, pa "S" u /miš ée/ postaje "S' ". Velarno /n/ u /banka/ bilježi se velikim slovom "N".

Naglasak se označava navodnicima ispred naglašenog sloga. Dugi akcent označava se dvotočkom iza vokala a silazni ili uzlazni ton označavaju se s F (falling) ili R (raising) zabilježeno u drugoj liniji ili odvojeno s <.>. Na primjer, /ruka/ treba transkribirati: "ru:ka<R>".

Taj je prijedlog recenzirao profesor Wells i može se neposrednim pristupom naći na adresi:

<http://www.phon.ucl.ac.uk/home/sampa/croatian.htm>

Za druge europske jezike popisi prekodiranih IPA simbola nalaze se na adresama kojima umjesto *croatian* treba upisati ime jezika (*english, bulgarian ...*)

Ovaj skup simbola nije zatvoren niti konačan. Moguće je unositi dodatke i alternativne prijedloge.

Osim SAMPA postoji i: SAMPROSA, sustav simbola za bilježenje prozodije, te SAMSINT, sustav simbola za bilježenja intonacije

Moguće kolizije između pojedinih zapisa (ako se koriste već upotrebljeni znakovi) izbjeci će se zapisivanjem na različitim linijama ili odvajateljima kao <>.

Autor SAMPA prijedloga Wells piše: SAMPA je kao i IPA skup, repertoar simbola primarno namijenjenih označavanju fonema, distinktivnih glasova nekog jezika a ne "hvatanju" svih alofonskih (subfonemičkih) detalja koji se mogu zapaziti ili na artikulacijskom ili akustičkom planu. Različiti fonetičari - istraživači na različite načine koriste predložene simbole. Tako, kada se engleski transkribira fonemski, zanemaruju se srazmjerno velike razlike u glasovima koje nastaju u različitim fonetskim kontekstima - /p t k/ može biti aspirirani i neaspirirani, /t/ može biti dentalni nasuprot alveolarnom itd. Bilo bi neekonomično sve te razlike zapisivati u svakoj transkripciji, kada se može umjesto toga navesti nekoliko fonoloških pravila. I SAMPA i IPA moguće je koristiti za bilježenje alofonskih detalja, patološkog izgovora, emfatičkog izgovora itd., ali to nije njihova primarna namjena. Osim toga, i SAMPA i IPA koriste iste simbole u različitim jezicima za označavanje glasova koji su objektivno međusobno različiti. Kada se piše /t/ u engleskom, on je u načelu apikoalveolaran i aspiriran dok isti simbol u francuskom označava glas koji je u načelu dentalan i neaspiriran, a u švedskom dentalan i aspiriran. Bilo bi beznadno neekonomično inzistirati na različitom simbolu za svaku različitost među glasovima svjetskih jezika.

Moramo biti svjesni toga da je komunikabilnost pomoću IPA i SAMPA vrlo uvjetovana. Naime, simbol i kratak artikulacijski opis, kako se navodi u popisima znakova, ni u kojem slučaju nisu dovoljni da prenesu informaciju o zvuku. Zbog toga se pribjegava navođenju primjera. No, primjeri vrijede samo za one korisnike koji srazmjerno dobro poznaju jezik u kojem se primjer navodi. SAMPA transkripcija korisna je već zato da se nedvosmisleno mogu prenijeti simboli fonemske transkripcije među govornicima istog jezika onda kada tradicionalna grafija nije dio ASCII repertoara.

Donosimo reprint stranica na adresi:

<http://www.phon.ucl.ac.uk/home/sampa/croatian.htm>

Namjerno je ostavljen originalan tekst napomene kako je u originalnoj verziji, bez hrvatskih grafema.

Croatian

Note: in the representation of Croatian orthography shown below,

* c' stands for c with an acute accent;

* d' stands for a barred d;

* c^ s^ z^ stand for c s z with a wedge accent (hacek).

SAMPA symbol Orthography Transcription Gloss

Vowels. There are five vowels in Croatian. Additionally, r can function as a vowel.

i	sir	"sir <F>	cheese
e	pet	"pe:t <F>	five
a	sat	"sa:t <F>	watch
o	bol	"bo:l <F>	pain
u	put	"pu:t <F>	way
r	prst	"prst <F>	finger

Consonants. The Croatian consonant system comprises 17 obstruents and 7 sonorant phonemes.

p	pas	"pas <F>	dog
b	bas	"bas <F>	base
t	tip	"ti:p <F>	type
d	dan	"da:n <F>	day
k	kap	"ka:p <F>	drop
g	gol	"go:l <F>	naked
ts	cura	"tsura <F>	girl
tS	c^ar	"tSa:r <F>	charm
dZ	dz^ep	"dZep <F>	pocket
tS'	noc'	"no:tS' <F>	night
dZ'	d'ak	"dZ'a:k <F>	pupil
f	fonetika	fo"netika <R>	phonetics
s	sat	"sa:t <F>	hour
z	zid	"zi:d <F>	wall
S	s^ef	"Se:f <F>	chief
Z	z^ena	"Zena <R>	woman

x	hod	"xo:d <F>	walking
m	most	"mo:st <F>	bridge
n	noga	"noga <R>	leg
J	njiva	"Jiva <F>	field
l	lud	"lu:d <F>	mad
L	ljut	"Lu:t <F>	angry
r	ruka	"ru:ka <R>	hand
v\	vid	"vi:d <F>	sight
j	ja	"ja: <F>	I

Conventions (Assimilation). Phonemes in contact can be assimilated and pronounced as allophonic variants. Some of these allophonic variants can be regarded as the neutral realizations of some other phoneme (for example, replacement of voiced obstruents by their voiceless counterparts and vice versa). Such cases are not exemplified here. We exemplify those allophones which are different from neutral realizations of Croatian phonemes.

//	[]	Orthog.	Transcript.	Gloss
f	v	s^ef bi	"Se:vbi	boss would
ts	dz	zec bi	"ze:dzbi	rabbit would
h	G	dah bi	"daGbi	breath would
S	S'	mis^ c'e	"miS'tS'e	mouse will
Z	Z'	ruz^ d'ine	"ru:Z'"dZ'ine	Gina's lipstick
m	F	tramvaj	"traFv\aj	tram
n	N	banka	"baNka	bank

Stress and Accent. There are two word tones in Croatian: -F, -R. Accented vowels can be short or long. The intersection of tones and lengths in accented syllables results in four word accents. They are not shown in orthography.

long fall	pas	"pa:s <F>	belt
short fall	pas	"pas <F>	dog
long rise	ruka	"ru:ka <R>	hand
short rise	noga	"noga <R>	leg

This page is maintained by J.C. Wells. It was created 1997 May 06 on the initiative of Juraj Bakran and Damir Horga.

REFERENCIJE

- Bakran, J., Horga, D.** (1997). Sampa for Croatian, Internet stranica na adresi:
<http://www.phon.ucl.ac.uk/home/sampa/croatian.htm>
- IPA** (1993). The International Phonetic Alphabet, *Journal of the IPA* 23:1.
- Wells, J., Barry, W., Grice, M., Fourcin, A. i Gibbon, D.** (1992). Standard Computer-Compatible Transcription. Esprit project 2589 (SAM), Doc. no. SAM-UCL-037. London: Phonetics and Linguistics Dept., UCL.

Juraj Bakran

Faculty of Philosophy, Zagreb, Croatia

Damir Horga

Faculty of Philosophy, Zagreb, Croatia

SAMPA FOR CROATIAN

SUMMARY

The article reports the proposition concerning a group of computer readable ASCII symbols that are expected to replace the standard IPA symbols for the phonetic transcription of Croatian speech sounds. The proposition has been based on SAMPA (Speech Assessment Methods Phonetic Alphabet) principles published by Wells et al. in 1992. The basic Croatian version is available on Internet and subject to modifications and additions.

Key words: *phonetic transcription, SAMPA, Croatian language*
