

FONETICKÁ TRANSKRIPCIA SAMPA A SLOVENČINA

Jozef Ivanecký – Mira Nábělková

1. ÚVOD

Príspevok predstavuje aplikáciu medzinárodného fonetického transkripčného systému SAMPA na slovenčinu ako aktuálne potrebný krok v jej sprístupnení pre počítačové spracovanie najmä v rámci viacjazyčných projektov. Impulzom pre vznik systému SAMPA a abecedy SAMPA bol rozmach v oblasti počítačového spracovania reči. Význam fonetickej transkripcie v počítačovom rozpoznávaní reči a jej syntéze nie je azda potrebné zdôrazňovať.

Systém na počítačové rozpoznávanie reči rozpoznáva na najnižšej úrovni jednotlivé zvuky, ktoré sú označované znakmi fonetickej abecedy zvolenej pre konkrétnu fonetickú transkripciu. Vo väčšine operačných systémov dodnes nie je uspokojivo vyriešená podpora pre prácu so symbolmi medzinárodného fonetického transkripčného systému IPA (API), preto sa na fonetický prepis začali využívať rôzne abecedy z počítačovo jednoducho kódovateľných štandardných ASCII znakov. Prirodzená nekompatibilita takýchto rozličných systémov prekážajúca vzájomnej komunikácii pri spracovaní jednotlivých jazykov viedla k projektu SAM (Speech Assessment Methods) (Fourcin – Harland – Barry – Hazan, 1989) a v jeho rámci k vytvoreniu medzinárodnej abecedy SAMPA cestou prevodu transkripčných znakov IPA do kódovania pomocou štandardných ASCII znakov.

S postupným rozširovaním znakov využívaných v rámci SAMPA abecedy rástol aj počet jazykov, na ktoré bola SAMPA aplikovaná a aplikovateľná. Podľa jej autorov je dnes plne kompatibilná s transkripčným systémom IPA. Táto postulovaná kompatibilita však ešte neznamená ľahkú aplikovateľnosť na slovenčinu. V ďalšom texte sa zaoberáme príčinou týchto problémov a ich rôznymi riešeniami od roku 1995 až po súčasný stav.

2. SAMPA – VÝVOJ A SÚČASNOSŤ

Hlavným dôvodom vzniku SAMPA systému nebol ani tak prudký vývoj v počítačovom spracovaní reči, ako skôr nárast prác v oblasti viacjazyčných systémov, kde sa vzájomná nekompatibilita pri rozličnom kódovaní použitom pre fonetickú transkripciu prejavila v plnom svetle.

Prvá verzia SAMPA fonetickej abecedy (PA) vznikla na konci deväťdesiatych rokov ako jeden z produktov európskeho projektu *Speech Assessment Methods* (SAM). Z týchto skratiek vznikol názov pre SAMPA abecedu. Hlavným koordinátorom prác na abecede SAMPA je od roku 1987 University College v Londýne.

Pri vytváraní tohto štandardného transkripčného systému aplikovateľného na mnohé európske jazyky sa vychádzalo z monojazyčných systémov uvedených v tabuľke č. 1. Rovnako veľký význam ako dosiahnutie vzájomnej kompatibility mala pri jeho vytváraní aj jednoduchosť kódovania pri počítačovom spracovaní.

Z tabuľky č. 1 je evidentné, že kódovacie tabuľky COST 209, Esprit 291 a Alvey majú veľa spoločného. Na základe toho boli postavené základy pre európsky štandard. Z princípov formulovaných v projekte Esprit 291 vyplýva nasledujúce: ASCII znaky v IPA tabuľke ostávajú nezmenené, nie ASCII znakom – do tej miery, do akej je to možné – sú priradené veľké ASCII znaky (napr.: a – A, e – E, ε – E).

Aplikovaním týchto pravidiel na tabuľku č. 1 dostaneme ASCII reprezentáciu pre nasledujúce znaky z fonetickej abecedy IPA:

IPA znaky	ɑ	ɔ	ɔ̃	ε	ə	ɤ	ɪ	ɨ	θ	ʃ	ʊ	ʏ	ʒ	ʔ	:
kódované															
ako ASCII	A	O	D	E	@	G	I	N	T	S	U	Y	Z	?	:

Na základe tohto a ďalších pravidiel bola formovaná prvá verzia SAMPA abecedy, prvýkrát publikovaná v (Wells, 1987). Táto fonetická abeceda, predstavujúca prvý – počítačovo jednoducho reprezentovateľný – kódovací systém dohodnutý medzi predstaviteľmi ôsmich európskych krajín spolupracujúcich na projekte SAM, je v tabuľke č. 2 a č. 3. Súčasťou tejto verzie sú aj znaky pre prozodické vlastnosti, ktoré boli neskôr vyčlenené do samostatného projektu SAMPROSA. Prvá verzia SAMPA svojím súborom transkripčných znakov pokrývala tieto jazyky: angličtina, dánčina, francúzština, nemčina, španielčina a taliančina.

Do projektu SAMPA nebol v čase jeho vzniku zahrnutý žiaden zo slovanských jazykov, z čoho vyplývala aj značná limitácia použitia abecedy SAMPA v prvej verzii pre slovanské jazyky. Napriek týmto obmedzeniam sme v roku 1995 vytvorili prvú neoficiálnu verziu SAMPA abecedy pre slovenčinu (Krokavec – Filasová – Ivanecký, 1996).

V posledných piatich rokoch bola SAMPA v rámci niekoľkých projektov značne rozšírená o nové transkripčné znaky a dnes už pokrýva mnohé – nie len indoeurópske – jazyky. Z priestorových dôvodov nie je možné, aby súčasná verzia SAMPA abecedy bola súčasťou tohto príspevku. Aktuálny stav SAMPA abecedy možno nájsť na (Wells, 1998).

Zo slovanských jazykov dnes existujú oficiálne verzie SAMPA abecedy pre bulharčinu, chorvátčinu, poľštinu, ruštinu a slovinčinu. Oficiálna verzia pre slovenčinu a ani pre češtinu dodnes nie je k dispozícii.

	IPA		COST	Esprit	Alvey	VL	BDSON	Allen
1	ɑ	!	A	A	A	A	a>	a (aa)
2	æ		{ 123	\	&			ae
3	ɐ		7	R \	6			
4	ɒ		C		Q			ao
5	ɔ	!	O	O	O	O	o>	O
6	ç		c	C	C			c-
7	ð	!	D	D	D			D
8	ɛ	!	E	E	E	E	e>	E
9	ə	!	@ 64	@	@	@	&	6
10	ɜ		3	@/	3			3
11	ɝ	!	G	G	G	G		g-
12	ɪ	!	I	I	I			I
13	ʎ		l,	L	F			lj
14	ɲ		n,	N	J		nj	nj
15	ŋ	!	N	N	N	N	ng	N
16	ø		124	q	{ 123	&	E	oe
17	œ		\92	Q	9	@	E>	oE
18	œ		&					OE
19	θ	!	T	T	T			T
20	ʃ	!	S	S	S	S	S	S
21	ʊ	!	U	U	U			U
22	ʌ		V	^	V			A
23	Y	!	Y	Y	Y			Y
24	ɥ			H	H		Y	
25	ʒ	!	Z	Z	Z	Z	Z	Z
26	?	!	?	?	?			?
27	ʁ,R		R	R	K		R	R
28			E ⁻	E ⁻	E+ ⁻		IN	E,n
29	aɪ		A>I	a/	aɪ	(Y)		ai
30	tʃ		I^S	T/	IS			
31	:	!	:	:	:			:
32	,		(')	,	"			"
33	,		(')	"	%			,

Tabuľka 1: Prehľad počítačovo reprezentovateľných fonetických transkripcií. Znak '!' v treťom stĺpci poukazuje na zhodu medzi tabuľkami COST 209, Esprit 291 a Alvey.

IPA	SAM-PA	ASCII	IPA	SAM-PA	ASCII
Samohlásky			Spoluhlásky		
a	a	97	b	b	98
ɑ	A	65	c	c	99
æ	{	123	ç	C	67
ɐ	6	54	d	d	100
ɒ	Q	81	ð	D	68
ɔ	O	79	f	f	102
e	e	101	g	g	103
ɛ	E	69	ɣ	G	71
ə	@	64	h	h	104
ɜ	3	51	j	j	106
i	i	105	k	k	107
ɪ	I	73	l	l	108
o	o	111	ʟ	L	76
ø	2	50	m	m	109
œ	9	57	n	n	110
Œ	&	38	ɲ	J	74
u	u	117	ŋ	N	78
ʊ	U	85	p	p	112
ʉ	}	125	r	r	114
ʌ	V	86	ʀ,R	R	82
y	y	121	s	s	115
Y	Y	89	ʃ	S	83
			t	t	116
			θ	T	84
			v	v	118
			w	w	119
			x	x	120
			ɸ	H	72
			z	z	122
			ʒ	Z	90
			ʻ	?	63

Tabuľka 2: SAM abeceda (SAMPA) pre fonetickú transkripciu – 1987.

IPA	SAM-PA	ASCII	Komentár
Dvojnakové symboly			
œ	9	57,126	Príklad nosovky
aɪ	aI	97,73	Príklad diftongu
tʃ	tS	116,83	Príklad afrikáty
ŋ	=n	61,110	Príklad slabičného /n/
	E/	69,47	Príklad archifonémy

IPA	SAM-PA	ASCII	Komentr
Hraničné a prozodické vlastnosti			
:		58	Znak pre kvantitu
”		34	Znak pre hlavný prízvuk
%		37	Znak pre vedľajší prízvuk
–		45	Znak pre úroveň tónu
’		39	Znak pre stúpavý tón
‘		96	Znak pre klesavý tón
“		96,39	Znak pre stúpavo-klesavý tón
”		39,96	Znak pre klesavo-stúpavý tón
\$		36	Znak pre slabičnú hranicu
		124	Znak pre slabičnú hranicu vnútri skupiny samohlások
-		45	Rozdeľovač

Tabuľka 3: SAM abeceda (SAMPA) pre fonetickú transkripciu – 1987.

3. SAMPA A SLOVENČINA

Prekódovanie existujúcej transkripcie do abecedy SAMPA bolo jednoduché v prípade jazykov, pri ktorých bola v podstate vyhovujúca a používaná transkripcia IPA, čo však nebol a stále nie je prípad slovenčiny. V slovenskom prostredí je na vedecké účely používaný domáci typ systému fonetickej transkripcie. V jeho vývine (a pre poznanie jeho vývinu) sa ako dôležitý javí štandardizačný návrh zo 60. rokov (Isačenko – Romportl, 1966). Pri revidovaní dovtedajších transkripčných prístupov sa vychádzalo z predstavy o funkčnosti paralelnej existencie istého počtu transkripčných sústav podľa potrieb konkrétnych úloh. Navrhovaná transkripcia mala predstavovať "stredný typ" fonetickej transkripcie medzi klasickou domácou dialektologickou transkripciou a medzinárodne uznanou transkripčnou sústavou (v danom prípade IPA), určený pre vedecké publikácie o češtine a slovenčine a ako základný typ aj pre učebnice cudzích jazykov. Zohľadňovalo sa v nej viacero faktorov, medziiným dovtedajší úzus v zapisovaní nárečových textov, ale aj dobové vybavenie tlačiarň a napokon aj písacích strojov. Vývin našej fonetickej transkripcie, konkrétne fakt, že sa výraznejšie v praxi napriek nesporným výhodám medzinárodne dohodnutého fonetického prepisu neujala transkripcia IPA, súvisel okrem špecifickosti IPA symbolov a technickej zložitosti pri ich tlačení aj s tým závažnejším problémom, že zostavovatelia transkripcie IPA sa opierali predovšetkým o poznanie neslovanských jazykov, s čím súviseli ťažkosti pri jej používaní na zápis prejavov v slovanských jazykoch podobne ako aj v mnohých orientálnych jazykoch (Isačenko – Romportl, 1966).

Keďže staršie ciele fonetického zápisu (prepisy slovenských najmä nárečových textov boli určené predovšetkým domácim bádatelom, sekudárne aj nelingvistickej verejnosti, z čoho v neposlednom rade vyplývala aj potreba čitateľnosti či čítavosti transkribovaného textu) ani nepredpokladali využívať prepis do znakov IPA, vystačilo sa s jednoduchším systémom založeným na grafickej sústave jazyka, v ktorom sa na prepis hlásky volí to písmeno, ktoré je v bežnej písanej forme jeho základným ekvivalentom, pravda niektoré znaky pre špecifické hlásky boli postupne do tohto systému doplnené. Vedomie dôležitosti medzinárodnej transkripcie IPA pritom vystupovalo aj za takýchto okolností a inšpirovalo ďalšie hľadania v tomto smere: "systém IPA je dnes tak rozšírený, že ho pri riešení otázok fonetickej transkripcie musíme brať do úvahy aj vtedy, keď chceme navrhnúť akúkoľvek inú všeobecnejšie zrozumiteľnú fonetickú transkripciu" (Kráľ, 1974b).

V tabuľke č. 4 je uvedený zoznam fón spisovnej slovenčiny kódovaných znakmi používanými v slovenskej fonetickej transkripcii podľa (Kráľ, 1983). Daná fonetická abeceda síce vyhovuje domácemu použitiu, nestačí však pri použití na medzinárodnom poli.

Využívanie domácej fonetickej transkripcie a nedoriešenosť jej vzťahu k transkripcii IPA trvá v slovenskom prostredí dodnes. Novšie je k dispozícii aj porovnávací prepis domáceho transkripčného systému a medzinárodnej transkripcie IPA (Dvončová a kol., 1988), no v oblasti slovenskej fonetickej transkripcie sa nevelmi

využíva a niektoré jeho riešenia možno považovať za diskusné.

Problémy, na ktoré sme narazili pri pokuse vytvoriť inventár slovenských fón kódovaných pomocou SAMPA fonetickej abecedy, môžeme zhrnúť do dvoch bodov:

- Nedoriešenosť vzťahu slovenskej fonetickej transkripcie s transkripciou IPA. Riešenie tohto problému bolo pre nás veľmi dôležité práve z toho dôvodu, že SAMPA abeceda je formovaná pomocou IPA transkripčného systému. Z toho vyplýva, že pri vytváraní SAMPA transkripcie pre nový jazyk stačí zobrať existujúci fonetický repertoár opísaný pomocou IPA systému a prepísať ho do SAMPA systému pomocou konverznej tabuľky SAMPA-IPA. Predpokladá to, pravda, existenciu prijatej IPA transkripcie pre konkrétny jazyk.
- SAMPA abeceda v čase svojho vzniku pokrývala len dosť malú časť repertoáru IPA. S určitými obmedzeniami a zavedením dvoch nových znakov dala sa použiť na vytvorenie fonematickej abecedy slovenčiny (Ivanecký, 1998), no vzhľadom na to, že neobsahovala znaky pre väčšinu slovenských variantov foném (porov. tab. č. 2), na vytvorenie fonetickej verzie vhodná nebola. Tento stav bolo možné zovšeobecniť pre väčšinu slovanských jazykov.

Pri vytváraní IPA verzie slovenskej transkripcie sme okrem neúplnosti zdrojov zaoberajúcich sa možnosťami aplikácie IPA systému na slovenčinu narazili aj na isté malé rozdiely v (Kráľ, 1983) a (Sabol, 1988) v opise slovenského transkripčného systému. Na naše účely sme ako východiskový zdroj použili (Kráľ, 1983). Tak bola už v roku 1995 na Katedre kybernetiky a umelej inteligencie Technickej univerzity v Košiciach vytvorená prvá verzia SAMPA abecedy pre slovenčinu. Vzhľadom na vyššie spomínané problémy šlo o fonematickú verziu. Keďže v tom čase SAMPA ešte neobsahovala znaky pre palatály *ť* a *ď*, priradili sme im dovedy nepoužité symboly. Táto verzia bola napriek obmedzeniam vyplývajúcim z absencie opisu alafón úspešne použitá v systéme na rozpoznávanie reči v rámci projektu SQEL (Aretoulaki a kol., 1998).

S rozširovaním SAMPA abecedy sa postupne vytvárali vhodné podmienky na vytvorenie slovenského SAMPA systému na fonetickej úrovni. Koncom 20. storočia bola SAMPA abeceda natoľko rozšírená o nové symboly, že už tento prechod umožňovala. Výsledkom našej doterajšej snahy je transkripčný systém SAMPA pre slovenčinu. Problematiku vypracovaného návrhu ďalej rozoberáme osobitne pre spoluhlásky a pre samohlásky.

Systémom SAMPA pre slovenčinu sa koncom 90. rokov pri riešení úloh spojených s počítačovou syntézou a analýzou zaoberalo aj iné pracovisko. V tejto verzii, okrajovo spomínanej v štúdiu z r. 1999 (Štefánik – Rusko – Považanec, 1999), vidíme isté nepresnosti.

Znak	Príklad	Znak	Príklad
<i>a</i>	papier	<i>r</i>	para, Andrej
<i>á</i>	pás	<i>ř</i>	prst, vrch
<i>ä</i>	pät, mäso	<i>ř</i>	vřba
<i>e</i>	pero	<i>l</i>	skala, vlak
<i>é</i>	dcéra nové	<i>l</i>	vlk
<i>i</i>	pivo, dym	<i>l</i>	vľča
<i>í</i>	pískať, nový	<i>ñ</i>	vaňa, niešť
<i>o</i>	popol	<i>l</i>	ľad, lipa, liek
<i>ó</i>	pól	<i>f</i>	fajka, vpiť, včela
<i>u</i>	puto	<i>v</i>	slovo, v oku
<i>ú</i>	púpava	<i>w</i>	vdova, vbiť
<i>ia</i>	piatok		v zime, šéflekár
<i>ie</i>	spievať	<i>ů</i>	pravda, otcov
<i>iu</i>	cudzíu		dlhou ulicou
<i>üo</i>	vôľa, kôň	<i>s</i>	osem, strom
<i>p</i>	popol	<i>z</i>	váza, zbierať
<i>b</i>	žaba	<i>š</i>	košela, prášiť
<i>t</i>	vata, tento	<i>ž</i>	veža, žaba
<i>d</i>	voda, dom	<i>x</i>	chata, pochvala
<i>ť</i>	Maťo, platiť	<i>h</i>	Praha, hrýzť
<i>ď</i>	háďa, deti	<i>γ</i>	vrch hory
<i>k</i>	páka, krk	<i>j</i>	dvaja, prijímať
<i>g</i>	gágať, agát	<i>ř</i>	kraj, krajský, zjest
<i>m</i>	mama	<i>c</i>	cena, pocta
<i>m</i>	amfiteáter, tamvon	<i>z</i>	hádzať, medza
<i>n</i>	rana, nos	<i>č</i>	oči, mačka
<i>n</i>	panský, inžinier	<i>ž</i>	džem, Hodža
<i>lj</i>	tenký, cengať		

Tabuľka 4: Tabuľka znakov slovenskej fonetickej transkripcie

4. SPOLUHLÁSKY

Pri prepise slovenských spoluhlások sa ako hlavný problém, s ktorým sa bolo v úvodných štádiách práce treba vyrovnáť, ukazovala neexistencia niektorých potrebných symbolov v raných verziách SAMPA.

Prepis časti spoluhlások bol jednoduchý tým, že napriek rozdielom slovenskej transkripcie a abecedy IPA sa v obidvoch systémoch určité percento spoluhlások prekrýva. Medzi zhodné znaky patria:

p, b, m, t, d, n, k.

Týmto znakom sme mohli priamo priradiť aj rovnaké SAMPA znaky, čo vyplýva z princípov definovaných pre vytváranie SAMPA transkripčného systému. Ďalšie znaky, ktorých prepis bol z rozličných dôvodov relatívne bezproblémový napriek tomu, že v IPA systéme sú pre ne použité odlišné symboly, sú:

g, r, l, ĺ, ň, v, f, g, x, s, z, š, ž, j, c, č, ʒ, ʝ.

Týmto bol repertoár slovenských spoluhlások (foném) prepísateľných pomocou prvej verzie SAMPA abecedy vyčerpaný.

Možno tu poznamenať, že pri niektorých z konsonantických IPA znakov sa v domácej diskusii (porov. Isačenko – Romportl, 1966) poukazovalo na istú problematickosť jednak z hľadiska ich "komunikačne neproblémovej" adekvátnosti, jednak z hľadiska paralelnej koexistencie transkripčných systémov – možný zdroj nedorozumení sa videl napr. v IPA znaku *c* pre palatálu *č* a v koexistencii domácej transkripcie a IPA okrem iného aj v tom, že IPA znak *ʒ* pre *ž* sa v domácej transkripcii využíval (a využíva) pre *dz*. Pri tvorbe abecedy SAMPA sme sa v konverzii prirodzene opierali o medzinárodne dohodnuté IPA znaky (priradenie IPA znakov k znakom slovenskej transkripcie vidno v tabuľke č. 5).

V ďalšom postupe sme palatálam *č*, *ď*, pre ktoré v abecede SAMPA v čase vzniku našej prvej verzie pre slovenčinu ešte neexistovali znaky, priradili vlastné symboly *ʎ* a *ɟ*, ktoré boli dovtedy nepoužité.

Z hľadiska vytvárania abecedy SAMPA pre slovenčinu boli najdôležitejšie nasledujúce rozšírenia SAMPA abecedy:

- **Zmena funkcie znaku ** – nealfabetické znaky v SAMPA abecede umiestnené za alfabetickým znakom menia niektorú kvalitatívnu vlastnosť predchádzajúcej fóny. Znak \ za znakom fóny z aktuálnej verzie znamená, že toto spojenie znakov reprezentuje novú fónu. Bol zavedený z dôvodu nedostatku voľných symbolov pre ďalšie roširovanie SAMPA abecedy.
- **Rozšírenie o gréčtinu**, ktoré obsahuje znaky pre už spomínané palatály *č* a *ď* (pre ktoré sme mali v prvej verzii pre slovenčinu vlastné symboly): pre *č*

SAMPA znak *c* zhodný s IPA znakom *c*, pre *ď* SAMPA znak *J* zodpovedajúci IPA znaku *j*.

Z týchto rozšírení SAMPA prepis pre slovenčinu získal symboly pre *ť* – *c*, *ď* – *J* a pre slovenskú znelú glotálu *h* – *h* (dovtedy sme ju prepisovali ako *h*).

V súvislosti s prepisom slovenských (aj českých) palatál možno poznamenať, že v aplikáciách rôznych transkripčných systémov sa na ich prepis niekedy nie vhodne využívajú znaky pre palatalizáciu. Bolo sa tomu treba vyhnúť aj v prepise do abecedy SAMPA a nepoužívať znak pre palatalizáciu zavedený skôr ako spomínané rozšírenie pre gréčtinu, ktoré prinieslo pre palatály *ť*, *ď* špeciálne znaky. Hoci slovenčina rovnako ako aj čeština má palatálne konsonanty a nemá palatalizované konsonanty, znak pre palatalizáciu sa aj v domácej transkripcii niekedy využíva napriek možnej zavádzajúcej interpretácii pri konfrontácii so skutočnými palatalizovanými konsonantmi, aké má napr. ruština – v medzinárodnom pohľade tak napr. konsonanty prepisované nie s mäkkčeňom, ale s horným alebo dolným apostroфом [dĕfi; d,et,i] vyznievajú ako palatalizované (používané znaky v domácej transkripcii pozri napr. u Dvončovej a kol., 1988, s. 28; o využití znaku pre palatalizovanosť v niektorých prepisoch češtiny do IPA porov. Palková, 1997, s. 39). V diskusii o "strednom type" domácej fonetickej transkripcie sa na daný fakt síce poukázalo, no jej navrhovatelia – upozorňujúc zase na iné možné problémy s využívaním symbolov IPA pre jednotlivé palatály a argumentujúc rozšírenosťou vedomia o tom, v ktorých prípadoch ide o palatálnosť a kde o palatalizovanosť – ostali pri tomto nejednoznačnom prepise (v návrhu bol prepis [d,et,i]) aj napriek vedomiu rozporu s požiadavkou po jednoznačnosti symbolov v medzinárodných podmienkach (Isačenko – Romportl, 1966).

Ďalší rozvoj abecedy SAMPA priniesol znaky aj pre jednotlivé alofóny slovenských spoluhlások (porov. Wells, 1998; znak *N* zodpovedajúci IPA znaku *ŋ* a *G* zodpovedajúci IPA znaku *ɣ* obsahovala už verzia abecedy SAMPA z r. 1987) a umožnil tak prepis slovenčiny na fonetickej úrovni.

V procese tvorby abecedy SAMPA pre slovenčinu sme v súvislosti s prepisom alofón *n* narazili na problém, ktorý sa doteraz nepodarilo celkom uspokojivo vyriešiť. Ide o prepis hlásky *n̥* (v použitej tabuľke Á. Kráľa reprezentuje tento znak frikatívnu alveolárnu apikálnu alofónu *n* v slovách ako *inžinier*, *slovenský*, inde však frikatívnu velárnu postdorzálnu alofónu *n*, napr. v slove *bronchitída* – tak napr. Kráľ – Sabol, 1989; Sabol, 1989). Vychádzame z predpokladu, že SAMPA podobne ako IPA (porov. aj Dvončová a kol., 1988) tu nepočíta s takým podrobným rozlíšením, keďže ani jeden z týchto systémov zodpovedajúci symbol nemá, a *n̥* prepisujeme ako *n* (napokon aj v existujúcich slovenských transkripčných systémoch existujú v zachytení alofón *n* popri kvalitatívnych aj kvantitatívne rozdiely).

5. SAMOHLÁSKY

Hlavným problémom pri vytváraní SAMPA abecedy pre slovenčinu boli samohlásky. V tomto prípade nešlo o nezrelosť SAMPA systému na opis samohlások, ale o rozchádzanie sa rôznych zdrojov na prepis slovenských samohlások do IPA systému, keďže slovenské samohlásky nie sú totožné ani s otvorenými ani so zavtorenými vokálmi v systéme IPA a každý prepis do IPA tak predstavuje vlastne len určité priblíženie k optimálnemu riešeniu na základe niektorých charakteristík samohlások.

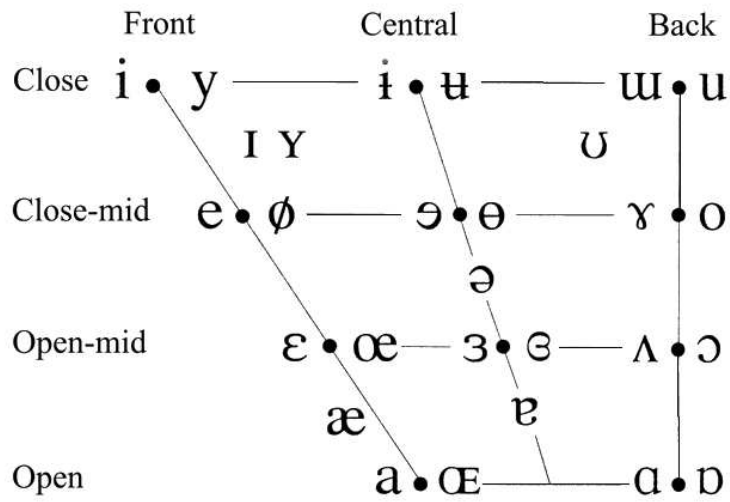
V prvej verzii abecedy SAMPA pre slovenčinu sme krátkym slovenským samohláskam **a, e, i, o, u** priradili IPA znaky **a, ɛ, ɪ, ɔ, ʊ** (Ivanecký, 1998).

V ďalšej verzii SAMPA (Ivanecký, 2001) sme v prepise slovenských samohlások uskutočnili zmenu a použili sme IPA znaky **a, ɛ, ɪ, o, u**. Pri tomto kroku sme sa opreli o novšie uvedenie IPA transkripcie pre češtinu, publikované v medzinárodnom zborníku IPA z roku 1999 (Dankovičová, 1999), pričom sme vychádzali z údajov fonetickej literatúry, že "slovenské spoluhlásky sa kryjú skoro úplne s českými, a to ako z hľadiska artikulácie, tak aj akustiky (Hála, 1960), (Hála, 1975), citované aj v (Dvončová – Jenča – Král, 1969). Iné práce (napr. Král, 1974b) však naznačujú, že postulovanie takýchto zhôd nemusí byť celkom presné. Na druhej strane aj česká fonetika stojí pri transkripcii do IPA pred problémom neoptimálnosti ponúkaných samohláskových symbolov, čo vedie aj v najnovších prácach k rozdielom v prepise krátkych vokálov: porov. napr. krátke **a, ɛ, ɪ, o, u** – dlhé **a:, ɛ:, i:, (o:), u:** (Dankovičová, 1999), krátke **a, ɛ, ɪ, ɔ, ʊ** – dlhé **a:, ɛ:, i:, o: u:** (Palková, 1997), krátke **a, ɛ, ɪ, o, u** – dlhé **a:, ɛ:, i:, o:, u:** (Encyklopedický slovník češtiny, 2002).

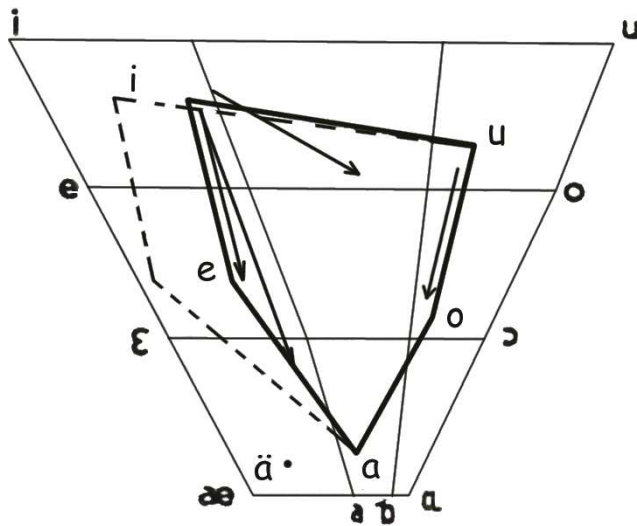
Pre konzistentné riešenie transkripcie slovenského vokalizmu bolo a je primerané upraviť prepis slovenských samohlások na základe ich uloženia v schéme kardinálnych vokálov (CV) podľa formantových frekvencií a priradenia najbližšieho IPA znaku. Pri hľadaní najprimeranejšieho znaku sa bolo treba prikloniť k istému riešeniu problému spomínanému v súvislosti s abecedou IPA už vyššie, totiž že slovenské samohlásky nie sú identické so samohláskami v systéme kardinálnych vokálov, ktorý sa využíva ako podklad a porovnávacie východisko pre rozličné vokalické systémy. Ich umiestnenie v schéme CV ukazujú tabuľky. Na obrázkoch č. 1 a č. 2 uvádzame vokalickú schému podľa rozpracovaného grafu IPA z r. 1996 (IPA, 1996) a uloženie slovenských vokálov, vychádzajúce z ich formantových charakteristík, podľa slovenskej fonetickej literatúry (Král – Sabol, 1987), porov. aj (Isačenko, 1968), (Král, 1974a), naznačujúce aj variačný rozptyl konkrétnych vokálov u rozličných osôb či v rozličnom hláskovom okolí.

Z uloženia slovenských krátkych samohlások v uvedenej schéme vyplýva, že využitie IPA symbolov **a, e, i, o, u** na ich prepis, ktoré nachádzme v niektorých prepisoch slovenských hlások do IPA aj SAMPA (Dvončová a kol., 1988), (Štefánik – Rusko – Považanec, 1999), posúva predstavu o ich vlastnostiach. Realite bližší je prepis pomocou znakov **a, ɛ, ɪ, ɔ** a **ʊ**, preto sme tieto znaky použili ako podklad

VOWELS



Obr. 1: Vokalická schéma IPA z roku 1996.



Obr. 2: Slovenské samohlásky a dvojhlásky v systéme kardinálnych vokálov. Čiar-
kovane je vyznačená poloha e, i po prepočte $(F2 + F3)/2$.

pre zápis v SAMPA.

Á. Král v analýze venovanej zápisu slovenských samohlások v znakoch IPA (Král, 1974a) pripomína väčšiu zatvorenosť slovenského *e* aj *o*, než je reprezentované týmto zápisom, čo naznačuje bodkou pod znakom, teda ε , \circ , keďže však pravidlá pre prepis kódovania IPA do SAMPA diakritiku tohto typu neobsahujú, použili sme v prepise ako východisko IPA znaky ε a \circ .

Pri prepise samohlásky *ä* sme využili IPA znak \ae (Á. Král v cit. článku vzhľadom na diftongoidný charakter samohlásky *ä* navrhoval ostať pri prepise pomocou jej vlastnej grafémy, čo sa však ťažko dá využiť pri kódovaní do SAMPA, napokon, tento znak je v IPA už aj obsadený).

V prepise dlhých samohlások sme na základe ich artikulačnej aj akustickej analógie s krátkymi samohláskami bez signifikantných rozdielov (Král, 1974a), (Král – Sabol, 1987) využili tie isté znaky ako pri krátkych, s označením dĺžky: **a:**, **ε:**, **I:**, **o:**, **U:**.

Pri prepise slovenských dvojhlasok sa ponúkajú v zásade dve možné riešenia. Jedným je využiť na prepis ich zložiek IPA resp. SAMPA znaky zodpovedajúce prepisu príslušných samohlások, a teda prepisovať ich ako: $\text{I}^{\wedge}\mathbf{a}$, $\text{I}^{\wedge}\mathbf{\varepsilon}$, $\text{I}^{\wedge}\mathbf{U}$, $\text{U}^{\wedge}\mathbf{O}$, resp. ($\text{I}^{\wedge}\mathbf{a}$, $\text{I}^{\wedge}\mathbf{E}$, $\text{I}^{\wedge}\mathbf{U}$, $\text{U}^{\wedge}\mathbf{O}$). Diskusia a alternatívne riešenia prepisu sa tu môžu odvíjať od realizačných špecifik slovenských dvojhlasok. Pri umiestnení do schémy CV na základe formantových charakteristík sa ukazuje, že ich krajné segmenty sa s príslušnými krátkymi samohláskami nezhodujú, pričom najväčší rozdiel možno vidieť v uložení druhého segmentu dvojhlasok *iu* v porovnaní so samohláskou *u* a druhého segmentu v *ia* v porovnaní so samohláskou *a*.

V prípade *iu* sme vzhľadom na centrálnejšie postavenie druhého segmentu využili symbol, ktorý na naznačenie centrálneho postavenia hlásky inak prepisovanej pomocou U ponúka SAMPA: $\text{U}\backslash$. Zodpovedajúci znak v IPA U uvádzame aj na mieste IPA symbolu v porovnávacíj tabuľke, dávame ho však do zátvoriek vzhľadom na to, že vo vokalických schémach IPA, ktoré máme k dispozícii, sa tento znak nevyužíva, a napokon sa na jeho predbežnú absenciu v IPA upozorňuje aj v prekódovacích pravidlách IPA–SAMPA (Wells, 1998).

Alternatívny zápis dvojhlasok *ia* ponúka (Král, 1974a), keď pre jej druhý segment, ktorý sa postavením blíži k postaveniu samohlásky *ä*, uvádza aj možnosť prepisu pomocou znakov \ae (SAMPA znak $\{$). Keďže však na základe artikulačno-percepčných vlastností má tento segment skôr a-ový charakter, prikláňame sa z daných alternatív prepisu dvojhlasok *ia* k možnosti $\text{I}^{\wedge}\mathbf{a}$ (SAMPA $\text{I}^{\wedge}\mathbf{a}$).

Navrhnutá transkripcia, ktorá sa opiera o doterajšie fonetické výskumy slovenských samohlások, o umiestnenie samohlások v CV a analýzu ich lokalizácie z hľadiska korešpondencie so symbolmi IPA aj v porovnaní s inými jazykmi (porov. (Král, 1974a), (Král, 1974b)), sa ukazuje ako možno nie definitívne a ideálne riešenie, ale ako schodná cesta pri vytváraní abecedy SAMPA.

Samozrejme, že aj v tomto prípade možno z hľadiska presnosti prepisu, teda priradenia zodpovedajúcich IPA symbolov ako podkladu pre prepis do SAMPA,

ŠT	IPA	SAMPA	ŠT	IPA	SAMPA	ŠT	IPA	SAMPA
a	a	a	p	p	p	ň	ɲ	J
á	a:	a:	b	b	b	ĺ	ʎ	L
ä	æ	{	t	t	t	f	f	f
e	ɛ	E	d	d	d	v	v	v
é	ɛ:	E:	č	c	c	w	f̥	f_v
i	ɪ	I	ď	ɟ	J\	ů	ʊ	U_ˆ
í	ɪ:	I:	k	k	k	s	s	s
o	ɔ	O	g	g	g	z	z	z
ó	ɔ:	O:	m	m	m	š	ʃ	S
u	ʊ	U	ɱ	ɱ	F	ž	ʒ	Z
ú	ʊ:	U:	n	n	n	x	x	x
ia	ɪa	L_ˆa	ɳ	-	n	h	ɦ	h\
ie	ɪɛ	L_ˆE	ŋ	ŋ	N	ɣ	ɣ	G
iu	ɪ(ʊ)	L_ˆU\	r	r	r	j	j	j
uo	ʊɔ	U_ˆO	ř	ʀ	r=	ĩ	ɪ	L_ˆ
			ṛ̌	ʀ:	r=:	c	ts	ts
			l	l	l	ʒ	dz	dz
			ḷ	l	l=	č	tʃ	tS
			ĺ	l:	l=:	ž	dʒ	dZ

Tabuľka 5: Porovnanie slovenskej transkripcie, IPA a SAMPA systému.

vzniesť námietky. SAMPA abeceda však bola navrhnutá pre počítačové spracovanie reči, čo znamená hlavne rozpoznávanie a syntézu, a z tohto pohľadu sa formantové frekvencie ukazujú ako cesta najvhodnejšieho priblíženia.

Ak sa na problematiku dívame z pohľadu počítačového rozpoznávania reči, takéto priblíženie v prepise SAMPA, korešpondujúcom zo zodpovedajúcimi symbolmi IPA, je vyhovujúce. V prípade monojazyčného systému bude príslušný symbol pokrývať priestor zodpovedajúci danému jazyku. V prípade viacjazyčného systému sa rozšíri daná oblasť a stredná hodnota sa posunie potrebným smerom. Keďže sa na rozpoznávanie dnes používajú výlučne štatistické metódy, môžeme o tom uvažovať ako o zväčšení rozptylu pre oblasť patriacu konkrétnym vokálom.

6. ZÁVER

Tabuľka č. 5 obsahuje našu verziu SAMPA pre slovenčinu. Okrem toho sú v nej uvedené aj príslušné IPA znaky.

Fonetická transkripcia SAMPA ako vo vzťahu k transkripcii IPA alternatívny, počítačovo jednoducho kódovateľný, a preto pre akékoľvek transkripčné ciele ľahšie využiteľný medzinárodný transkripčný systém, môže nájsť uplatnenie aj mimo

oblasti počítačového spracovania reči. SAMPA, ako sme sa v príspevku usilovali ukázať, zaznamenala v rokoch svojho vývinu zmeny, ktoré ju v súčasnosti robia kompatibilnou a súbežne použiteľnou s transkripciou IPA pri prepise mnohých jazykov, a až na spomínané alofóny *n* umožňuje dnešný "medzinárodne zaraditeľný" fonetický prepis slovenčiny.

Jozef Ivanecký, Katedra kybernetiky a umelej inteligencie Technickej Univerzity Košice, Švermova 9, 040 01 Košice; IBM European Speech Research, Vangerowstrasse 18, 69115 Heidelberg, Nemecko

Mira Nábělková, Jazykovedný ústav Ľudovíta Štúra, Panská 27, 851 04 Bratislava

Bibliografia

ARETOULAKI M. – HARBECK S. – GALLWITZ F. – NOETH E. – NIEMANN H. – IVANECKY J. – IPSIC I. – PAVESIC N. – MATOUSEK V.: *SQEL: A Multilingual and Multifunctional Dialogue System*. In: Proc. Int. Conf. on Spoken Language Processing, Sydney, Australia, December 1998

FOURCIN A. – HARLAND G. – BARRY W. – HAZAN V.: *Speech input and output assessment-multilingual methods and standards*. Ellis Horwood Limited 1989.

DANKOVIČOVÁ D.: *Czech IPA*. In: Handbook of the International Phonetic Association, Cambridge University Press 1999, s. 70–73.

DVONČOVÁ J. a kol.: *Fonetika a fonológia. Otázky transkripcie*. Bratislava, Univerzita Komenského v Bratislave 1988. 180 s.

DVONČOVÁ J. – JENČA G. – KRÁL A.: *Atlas slovenských hlások*. Bratislava, SAV 1969. 105 s.

Encyklopedický slovník češtiny. Ed. P. Karlík – M. Nekula – J. Pleskalová. Praha, Nakladatelství Lidové noviny 2002.

HÁLA B.: *Fonetické obrazy hlásek*. Praha, Státní pedagogické nakladatelství 1960. 127 s. + 43 s. obrazová příloha.

HÁLA B.: *Fonetika v teorii a v praxi*. Praha, SPN 1975.

The International Phonetic Association: *Reproduction of The International Phonetic Alphabet*. URL: <http://www2.arts.gla.ac.uk/IPA/fullchart.html>

ISAČENKO A. V.: *Spektrografická analýza slovenských hlások*. Bratislava, SAV 1968.

- ISAČENKO A. V. – ROMPORTL M.: *Návrh fonetické a fonologické transkripcie češtiny a slovenštiny*. Slovo a slovesnosť, 27, 1966, s. 163–166.
- IVANECKÝ J.: *Automatizácia fonematickej transkripcie slovenčiny*. In: Varia VII. Ed. M. Nábělková. Bratislava, SJS pri SAV 1998, s. 48–52.
- IVANECKÝ J.: *SAMPA v slovenčine a jej význam z pohľadu viacjazyčných systémov na rozpoznávanie reči*. In: Slovko. Slovenčina a čeština v počítačovom spracovaní. Ed. A. Jarošová. Bratislava, Veda 2001, s. 98–108.
- KRÁĽ Á.: *Fonetická povaha a systém slovenských samohlások a dvojhlások*. In: Studia Academica Slovaca. 3. Ed. J. Mistrík. Bratislava 1974a, s. 51 – 66.
- KRÁĽ Á.: *Pravidlá slovenskej výslovnosti*. Slovenské pedagogické nakladateľstvo, Bratislava 1983. 632 s.
- KRÁĽ Á.: *Slovenské samohlásky v konfrontácii so samohláskami iných jazykov*. In: Studia Academica Slovaca 3. Ed. J. Mistrík. Bratislava 1974b, s. 67–82.
- KRÁĽ Á. – SABOL J.: *Fonetika a fonológia*. Slovenské pedagogické nakladateľstvo, Bratislava 1987. 392 s.
- KROKAVEC D. – FILASOVÁ A. – IVANECKÝ J.: *Basic Properties of the Phonetic Transcription and the Sentence Description in Slovak Language*. In: "COST-249 Meeting". Košice, FEI Technical University 1996.
- PALKOVÁ Z.: *Fonetika a fonologie češtiny*. Praha, Karolinum 1997. 366 s.
- PAULINY E.: *Slovenská fonológia*. Bratislava, Slovenské pedagogické nakladateľstvo 1979. 213 s.
- SABOL J.: *Syntetická fonologická teória*. Jazykovedný Ústav Ľ. Štúra SAV, Bratislava 1988. 255 s.
- ŠTEFÁNIK J. – RUSKO M. – POVAŽANEC D.: *Frekvencia slov, grafém, hlások a ďalších elementov slovenského jazyka*. Jazykovedný časopis, 50, 1999, s.
- WELLS J. C.: *Computer-coded phonetic transcription*. Journal of the International Phonetic Association 17, No. 2, s. 94–114, 1987.
- WELLS J. C.: *Computer-coding the IPA: a proposed extension of SAMPA*. 1998 URL: <http://www.phon.ucl.ac.uk/home/sampa/home.htm>, SAMPA – official WWW pages.